Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **500,** 5195–5213 (2021) doi:10.1093/mnras/staa3293
Advance Access publication 2020 October 23

DAYENU: a simple filter of smooth foregrounds for intensity mapping power spectra

Aaron Ewall-Wice [1,2]★ Nicholas Kern,[1] Joshua S. Dillon [1]† Adrian Liu,[3] Aaron Parsons,[1] Saurabh Singh,[3] Adam Lanman [4] Paul La Plante,[1,2] Nicolas Fagnoni [5] Eloy de Lera Acedo [5] David R. DeBoer,[1] Chuneeta Nunhokee,[1] Philip Bull [6] Tzu-Ching Chang,[7,8] T. Joseph W. Lazio,[7] James Aguirre[9] and Sean Weinberg[10]

[1]*Department of Astronomy, University of California, Berkeley, CA 94720, USA*
[2]*Berkeley Center for Cosmological Physics, University of California, Berkeley, CA 94720, USA*
[3]*Department of Physics and McGill Space Institute, McGill University, 3600 University Street, Montreal, QC H3A 2T8, Canada*
[4]*Department of Physics, Brown University, Providence, RI 02906, USA*
[5]*Cavendish Astrophysics, University of Cambridge, Cambridge CB2 1TN, UK*
[6]*School of Physics & Astronomy, Queen Mary University of London, London E1 4NS, UK*
[7]*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr, M/S 169-237, Pasadena, CA 91109, USA*
[8]*California Institute of Technology, 1200 E California Blvd, Pasadena, CA 91125, USA*
[9]*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*
[10]*QC Ware, Palo Alto, CA 94301, USA*

Accepted 2020 October 20. Received 2020 October 5; in original form 2020 May 5

ABSTRACT

We introduce DPSS Approximate lazY filtEriNg of foregroUnds (DAYENU), a linear, spectral filter for H I intensity mapping that achieves the desirable foreground mitigation and error minimization properties of inverse co-variance weighting with minimal modelling of the underlying data. Beyond 21-cm power-spectrum estimation, our filter is suitable for any analysis where high dynamic-range removal of spectrally smooth foregrounds in irregularly (or regularly) sampled data is required, something required by many other intensity mapping techniques. Our filtering matrix is diagonalized by Discrete Prolate Spheroidal Sequences which are an optimal basis to model band-limited foregrounds in 21-cm intensity mapping experiments in the sense that they maximally concentrate power within a finite region of Fourier space. We show that DAYENU enables the access of large-scale line-of-sight modes that are inaccessible to tapered discrete Fourier transform estimators. Since these modes have the largest SNRs, DAYENU significantly increases the sensitivity of 21-cm analyses over tapered Fourier transforms. Slight modifications allow us to use DAYENU as a linear replacement for iterative delay CLEANing (DAYENUREST). We refer readers to the Code section at the end of this paper for links to examples and code.

Key words: methods: data analysis – techniques: interferometric – techniques: spectroscopic – dark ages, reionization, first stars – large-scale structure of the Universe.

1 INTRODUCTION

Buried under vastly brighter foregrounds, redshifted 21-cm emission from H I at redshifts $z \gtrsim 6$ remains an elusive treasure trove of information on how the first stars and galaxies heated and subsequently ionized the Universe. Experiments seeking to observe spatial 21-cm fluctuations are attempting a first detection with the power-spectrum statistic, $P(k)$ defined through

$$(2\pi)^3\delta^D(\mathbf{k}-\mathbf{k}')P(\mathbf{k}) = \langle\widetilde{T}_b(\mathbf{k})\widetilde{T}_b^*(\mathbf{k}')\rangle - \langle\widetilde{T}_b(\mathbf{k})\rangle\langle\widetilde{T}_b^*(\mathbf{k}')\rangle, \quad (1)$$

where δ^D is the Dirac delta function, $T(\mathbf{k})$ is the co-moving spatial Fourier transform of the cosmological brightness temperature field

$$\widetilde{T}_b(\mathbf{k}) = \int \mathrm{d}^3\mathbf{r}e^{i\mathbf{k}\cdot\mathbf{r}}T_b(\mathbf{r}), \quad (2)$$

and $\langle\cdot\rangle$ denotes an ensemble average. Gaussian random fields are completely described by the power spectrum. The power spectrum is also a convenient statistic for non-Gaussian fields since we can take advantage of the fact that cosmological quantities approximately obey statistical homogeneity and isotropy; allowing us to build sensitivity by averaging in spherical Fourier bins.

Another convenient feature 21-cm and other intensity mapping experiments is that foregrounds; which are expected to be intrinsically spectrally smooth, only occupy small wavenumbers along the line of sight (LoS, small $k_{\parallel}$) while 21 cm and other spectral lines that trace cosmological structures have substantial fine-scale spectral features

★ E-mail: aaronew@berkeley.edu
† National Science Foundational Astronomy and Astrophysics Postdoctoral Fellow.

(Di Matteo, Ciardi & Miniati 2004; Datta, Bowman & Carilli 2010; Parsons et al. 2012b). Thus, the native Fourier space of the power spectrum is well-suited for performing foreground separation.

While single-dish experiments such as GBT have been used to detect the 21-cm power spectrum at low redshifts (Chang et al. 2010; Masui et al. 2013; Switzer et al. 2013; Anderson et al. 2018), many have been turning to interferometers for obtaining the necessary high sensitivities for detecting 21 cm at higher redshifts. Interferometric experiments seeking to detect 21-cm fluctuations include CHIME (Bandura et al. 2014), Tianlai (Chen 2015), Ooty (Subrahmanya, Manoharan & Chengalur 2017), HIRAX (Newburgh et al. 2016), the MWA (Tingay et al. 2013), LOFAR (van Haarlem et al. 2013), the LWA (Ellingson et al. 2009), and HERA (DeBoer et al. 2017). Interferometric data sets consist of cross-correlations (visibilities) measured by pairs of antennas (baselines) at various spectral frequencies. Since line-emission at different distance along the LoS ($r_\parallel$) is redshifted to different observed frequencies, one can map observed frequencies to co-moving distance $\nu \propto x_\parallel$. For a given visibility, the Fourier dual of frequency is the delay, τ between signals arriving at each antenna. Thus $\tau \approx 2\pi Y^{-1} k_\parallel$, where Y is a constant. We refer the readers to Morales & Hewitt (2004) and Parsons et al. (2012a) for the full expression. Smooth structures, such as foregrounds, reside at delays smaller then light traveltime between the two antennas, τ_H; a phenomena known as the 'wedge' (Datta et al. 2010; Morales et al. 2012; Vedantham, Udaya Shankar & Subrahmanyan 2012; Parsons et al. 2012b; Pober et al. 2013). The fine-scale 21-cm fluctuations reside at all delays. A natural analysis choice that has been adopted by most Cosmic Dawn fluctuations experiments is to estimate power spectra by applying a discrete Fourier transform (DFT) either on raw interferometric visibilities (Parsons et al. 2012b, 2014; Ali et al. 2015) or on gridded u–v data and/or images (Chapman et al. 2012; Dillon, Liu & Tegmark 2013; Dillon et al. 2015; Jacobs et al. 2016; Trott et al. 2016; Barry et al. 2019) and then squaring. In taking an unpadded DFT along a single axis (we consider the $r_\parallel$ axis for example) one replaces the integral in equation (2) with a discrete sum over N_d sampled data points.

$$\int \mathrm{d}r_\parallel \mathrm{e}^{-ik_\parallel^n r_\parallel} \rightarrow \Delta r_\parallel \sum_{m=0}^{N_\mathrm{d}-1} \mathrm{e}^{-ink_\parallel^n \Delta r_\parallel}, \tag{3}$$

where $\Delta r_\parallel$ is the interval between LoS samples and $k_\parallel^m$ is the nth discrete wavenumber, $k_\parallel^n = 2\pi n (N_\mathrm{d}\Delta r_\parallel)^{-1}$, $n \in \{0, \ldots, N_\mathrm{d} - 1\}$. Since foregrounds are confined to the wedge, these techniques can contain/avoid foregrounds by throwing away/downweighting visibility DFT modes with $\tau \lesssim \tau_\mathrm{H}$.

Two realities complicate DFT techniques, both of which are related to incomplete sampling. First, data are sampled over a finite bandwidth with a sharp cut-off at the band edges. Secondly, flagging (excising) of radio frequency interference (RFI) introduces gaps in frequency sampling with additional sharp edges. The DFTs of incompletely sampled foregrounds have (spectral) sidelobes that often greatly exceed the expected amplitude of the 21-cm signal.

A number of approaches have been adopted to overcome incomplete data coverage. Most address the problem of finite bandwidth by multiplying data by a tapering function that goes to zero at the band-edges (Thyagarajan et al. 2016; Kolopanis et al. 2019). These multiplicative tapering or apodization filters smoothly filter the components of the signal at the band edges that is affected by sharp finite sampling features. While this leads to signal loss, bringing the foregrounds gradually to zero near the band edges compactifies their footprint in the DFT basis. A number of techniques also exist to deal with flagged channels. Per-baseline delay CLEANing[1] (Parsons et al. 2012b) iteratively peels and fits foregrounds on each baseline with a limited number of smooth discrete Fourier modes, interpolating over the channel gaps. Rather than interpolating with DFT modes, FASTICA (Chapman et al. 2012) fits smooth independent components at each LoS in a data cube, and subtracts them before performing the DFT into bandpower space. εPPSILON (Barry et al. 2019), similar CLEAN, interpolates over channel gaps with a DFT eigenbasis via the Lomb–Scargle method (Lomb 1976; Scargle 1982). Unlike CLEAN, it also attempts to interpolate the 21-cm signal by fitting all DFT modes rather than modes within a low delay window.

Any power-spectrum method involves linear filtering, transforming into a power bandpower basis, squaring, and then normalizing squared band-powers with a linear operator can be described in the quadratic estimator (QE) formalism, including several of the already mentioned techniques. For example, while FASTICA iteratively determines a foreground subtraction matrix from the data, the application of this subtraction matrix to data can be cast as a QE. Tegmark (1997) showed that the optimal (information preserving and minimizing error bars) quadratic estimator (OQE) for the component of a Gaussian signal $\mathbf{x}$, that is completely described by discrete bandpowers, p^α is given by a QE where (1) the linear filter is the inverse of the data covariance $\mathbf{C}^{-1}$, (2) the transforming and squaring step is performed by the derivative of the total covariance with respect to each αth bandpower $\mathbf{C}_{,\alpha}$, and (3) the normalization matrix is equal to the inverse of the diagonal of the Fisher information matrix $\mathrm{Diag}(\mathbf{F})^{-1}$.

While this recipe is straightforward, several issues complicate its implementation. Perhaps most glaring is the fact that $\mathbf{C}$ not actually known to much precision. The low-level component from the 21-cm signal itself is completely unknown while our ability to characterize our instrument (Pober et al. 2012; Neben et al. 2015, 2016; Jacobs et al. 2017; Fagnoni et al. 2019) and low-frequency foregrounds (Jacobs et al. 2011; Carroll et al. 2016; Line et al. 2017; Zheng et al. 2017; Eastwood et al. 2018) is currently limited to the $\sim$1 per cent level.

This has led to attempts at estimating $\mathbf{C}$ directly from data (Ali et al. 2015; Dillon et al. 2015) and/or modelling it given our understanding of the foregrounds and instrument (Dillon et al. 2013; Shaw et al. 2014; Trott et al. 2016). Recent investigations have found that data-driven approaches run a high risk of unintentional signal loss (attenuation of the 21-cm signal) (Switzer et al. 2015; Patil et al. 2016; Cheng et al. 2018) which, if not corrected, led to highly biased results. Along the same vein, it is unclear how well model driven covariances must accurately represent the underlying data in order to be effective and whether inaccurate model co-variances face similar signal loss issues associated data derived co-variances.

Liu & Shaw (2019) point out that attenuation of cosmological modes does not necessarily constitute signal loss as long as we characterize and correct this attenuation downstream. Indeed, standard normalization choices in the literature are explicitly calculated to undo filtering biases. However, great care must be exercised. The assumptions under-girding normalization formulas are (as we shall see) easily violated.

Normalization matrices are also chosen to 'demix' the smearing between various bandpowers that arise from the non-identity transfer function of our experiment and data-reduction choices. Effective foreground filters introduce signal loss to foregrounds but not the 21-

[1]This method applies the CLEAN algorithm used in radio astronomy imaging (Högbom 1974) to one spectral dimension.

cm signal. Since filtering can introduce 21-cm signal loss, it is useful to determine whether and when one can abandon filtering altogether and mitigate all foreground leakage at the demixing normalization step after bandpowers have been formed.

This paper is part one of a two part series. In it, we demonstrate the existence of a simple, fast, and effective foreground filter that is capable of imparting large amounts of good signal loss on arbitrarily sampled spectrally smooth foregrounds. We examine the properties of this filter compare its performance to the traditional approach of band-power estimation with a windowed DFT. In paper two, we will carefully examine the requirements for successfully demixing and reversing signal loss in the normalization step along with the consequences of violating these requirements.

Our filter is based on a simple, analytical model for $\mathbf{C}$ which captures the essential features of foregrounds: that they are overwhelming bright compared to the signal, that they occupy a continuum of delays up to some maximum, and that we measure them at a finite number of band-limited frequencies. The computation of this covariance matrix can be performed very quickly, using simple closed-form expressions while its analytical simplicity also allows us to study the origins of its efficacy. Because our filter is diagonalized, under certain circumstances, by Discrete Prolate Spheroidal Sequences (DPSS; Slepian 1978), we call our method DPSS Approximate lazY filtEriNg of foregroUnds (DAYENU).[2] While we discuss DAYENU in the context of foreground filtering and power-spectrum estimation for 21-cm cosmology, DAYENU can be applied to intensity mapping with other lines (e.g. C II, CO, Ly α) where foreground are distinguished from cosmological fluctuations on the basis of spectral smoothness.

Our paper is organized as follows. In Section 2, we review the mathematical formalism for QEs. In Section 3, we introduce our simplified inverse covariance weighting scheme, studying its performance on idealized data, its signal loss properties, and its relationship to DFT filtering. In Section 4, we examine DAYENU's performance in foreground filtering and power-spectrum estimation with realistic simulations of foregrounds and 21-cm fluctuations observed by the Hydrogen Epoch of Reionization Array (HERA; DeBoer et al. 2017).

2 FORMALISM

In this section, we set up our notation and review the formalism of QEs and OQEs.

2.1 Bandpowers

The data $\mathbf{x}$ observed in a fluctuation experiment can be decomposed into foregrounds ($\boldsymbol{f}$), noise ($\boldsymbol{n}$), and cosmological fluctuations ($\boldsymbol{s}$)

$$\mathbf{x} = \boldsymbol{f} + \boldsymbol{n} + \boldsymbol{s}. \tag{4}$$

Since $\boldsymbol{f}$, $\boldsymbol{n}$, and $\boldsymbol{s}$ are independent, $\mathbf{C} = \langle \mathbf{x}\mathbf{x}^\dagger \rangle - \langle \mathbf{x} \rangle \langle \mathbf{x}^\dagger \rangle$ can be decomposed into

$$\mathbf{C} = \mathbf{C}_{\rm fg} + \mathbf{N} + \mathbf{S}, \tag{5}$$

where $\mathbf{N} = \langle \boldsymbol{n}\boldsymbol{n}^\dagger \rangle$, $\mathbf{S} = \langle \boldsymbol{s}\boldsymbol{s}^\dagger \rangle - \langle \boldsymbol{s} \rangle \langle \boldsymbol{s}^\dagger \rangle$, and $\mathbf{C}_{\rm fg} = \langle \boldsymbol{f}\boldsymbol{f}^\dagger \rangle - \langle \boldsymbol{f} \rangle \langle \boldsymbol{f}^\dagger \rangle$.

[2]In Hebrew, 'day' translates approximately to 'sufficient' and 'enu' means 'to us'. The acronym refers to the fact that our filter is sufficient to us for removing foregrounds for 21 cm and other intensity mapping data sets.

Bandpowers are usually defined by decomposing $\mathbf{S}$ into a set of response matrices

$$\mathbf{S} = \sum_\alpha p^\alpha \mathbf{C}_{,\alpha}. \tag{6}$$

While many authors stick with bandpowers that only describe $\mathbf{S}$, Parsons et al. (2014), Ali et al. (2015), and Liu, Parsons & Trott (2014a, b) adopt bandpower definitions where $\mathbf{C}_{\rm fg} + \mathbf{S} = \sum_\alpha p^\alpha \mathbf{C}_{,\alpha}$. The decision to define bandpowers for the signal covariance $\mathbf{S}$ alone versus $\mathbf{C}_{\rm fg} + \mathbf{S}$ is an analysis choice with important consequences that we explore in paper II. Since we do not know the 21-cm signal a priori, we do not actually know what the correct bandpowers to use are. Instead, we choose a set of response matrices $\widehat{\mathbf{C}}_{,\alpha}$ that may not actually be correct. A standard choice for $\widehat{\mathbf{C}}_{,\alpha}$ uses our expectation that the 21-cm signal is homogenous so that the correlation between temperatures at two locations is given by the continuous Fourier transform of the power spectrum. Authors usually replace this continuous Fourier Transform with a DFT. Thus, many works (e.g. Dillon et al. 2015; Trott et al. 2016; Barry et al. 2019; Mertens et al. 2020) choose $\widehat{\mathbf{C}}_{,\alpha} = \mathbf{C}_{,\alpha}{}^{\rm DFT}$. For a 3D data cube, each data point x_m has an associated co-moving position $\mathbf{r_m}$ so

$$\left[\widehat{\mathbf{C}}_{,\alpha}{}^{\rm DFT,\,3D}\right]_{mn} \propto \sum_{\mathbf{k}\in V_\alpha} {\rm e}^{-i\mathbf{k}\cdot(\mathbf{r}_m - \mathbf{r}_n)}, \tag{7}$$

where V_α are Fourier-space bins (cylindrical or spherical) and $\mathbf{k}$ are wavenumbers given by the DFT of a gridded image.

In this work, we focus on per-baseline QEs employed by PAPER and HERA (Parsons et al. 2012b, 2014; Ali et al. 2015) which operate independently on different baselines at different LSTs. These estimators sacrifice a small amount of sensitivity for short baselines (Zhang, Liu & Parsons 2018) and have the advantage of being analytically and computationally simple to work with. For a per-baseline estimator, $\mathbf{x}$ is the frequency data from a single visibility at a single LST that has potentially been averaged over many identical copies in a redundant baseline group and many different nights at the same LST. We emphasize that this estimator is distinctive from a multibaseline estimator where the data are $\mathbf{x}$ consists of all baselines in our data set (e.g. Liu et al. 2014a, b). The DFT bandpowers used in per-baseline estimators are usually just the squared coefficients of a 1D frequency DFT. If the baselines are all sufficiently close together, each spherical k-bin is the same as each $k_\parallel$ bin in the LoS DFT. Parsons et al. (2014), Ali et al. (2015), and in this paper, we focus on LoS DFT bandpowers

$$\left[\widehat{\mathbf{C}}_{,\alpha}{}^{\rm DFT}\right]_{mn} \propto {\rm e}^{-2\pi i m n/N_{\rm d}}. \tag{8}$$

2.2 Quadratic estimators

In the QE formalism, we denote our $N_{\rm b}$ estimates of bandpowers $\hat{p}_\alpha$ to be equal to a normalized linear combination pairwise multiplications of data points

$$\hat{p}_\alpha = \frac{1}{2}\sum_\beta M_{\alpha\beta}\mathbf{x}^\dagger \mathbf{E}_\beta \mathbf{x} - \hat{b}_\alpha, \tag{9}$$

where $\mathbf{E}_\beta$ is one of $N_{\rm b}$ different $N_{\rm d} \times N_{\rm d}$ matrices (one for each bandpower) that perform a weighted sum over pairs of data measurements. $\mathbf{M}$ is an $N_{\rm b} \times N_{\rm b}$ normalization matrix and $\hat{b}_\alpha$ is a subtracted estimate of the true bias b_α which includes all covariance contributions not described by bandpowers.

$$b_\alpha = \sum_\beta M_{\alpha\beta} {\rm tr}\left[\mathbf{E}_\beta \left(\mathbf{C} - \sum_\gamma \mathbf{C}_{,\gamma}\right)\right]. \tag{10}$$

It is convenient to expand $\mathbf{E}_\alpha$ into a product of filter matrices, $\mathbf{R}$, and a quadratic matrix, $\mathbf{Q}_\alpha$

$$\mathbf{E}_\alpha = \mathbf{R}^\dagger \mathbf{Q}_\alpha \mathbf{R}. \tag{11}$$

Under this expansion, $\mathbf{R}$ describes all filtering applied to data prior to Fourier transforming. For a single visibility, this could be the apodization by a Blackman–Harris window in which case $R^{\rm BH}_{mn} \equiv \delta^k_{mn} T^{\rm BH}_n$, where $\boldsymbol{\delta}^{\boldsymbol{k}}$ is the Kronecker delta matrix and $T^{\rm BH}_n$ is the nth element of a Blackman–Harris window. Alternatively, for inverse covariance weighting, we might set $\mathbf{R}^{\rm OQE} \equiv \mathbf{C}^{-1}$. $\mathbf{Q}_\alpha$ performs the transformation into the bandpower basis for both data vectors along with binning and squaring. A standard example for $\mathbf{Q}_\alpha$ used to estimate DFT bandpowers is the per-baseline delay-transform matrix

$$\left[\mathbf{Q}_\alpha{}^{\rm DFT}\right]_{mn} = {\rm e}^{-2\pi i \alpha (m-n)/N_{\rm d}}. \tag{12}$$

$\mathbf{M}$ is usually chosen in a way that trades off mixing between bandpowers and their error correlations. The expectation value of each estimated bandpower, $\hat{p}^\alpha$ is equal to an admixture of true bandpowers

$$\langle \hat{p}_\alpha \rangle = \sum_\beta W_{\alpha\beta} p_\beta + b_\alpha - \hat{b}_\alpha \tag{13}$$

where

$$\mathbf{W} = \mathbf{MH} \tag{14}$$

and

$$H_{\alpha\beta} = \frac{1}{2}{\rm tr}(\mathbf{R}^\dagger \mathbf{Q}_\alpha \mathbf{R} \mathbf{C}_{,\boldsymbol{\beta}}). \tag{15}$$

2.3 Optimal quadratic estimators

The OQE that minimizes error bars and preserves all information from the original data is given by (Tegmark 1997; Liu & Tegmark 2011)

$$\hat{p}^\alpha_{\rm OQE} = \left[{\rm Diag}(\mathbf{F})\right]^{-1}_{\alpha\alpha} \left[(\mathbf{C}^{-1}\mathbf{x})^\dagger \mathbf{C}_{,\alpha} (\mathbf{C}^{-1}\mathbf{x})\right] - b_\alpha, \tag{16}$$

where Diag($\mathbf{F}$) is the diagonal of the Fisher information matrix given by

$$F_{\alpha\beta} = \frac{1}{2}{\rm tr}\left[\mathbf{C}^{-1}\mathbf{C}_{,\alpha}\mathbf{C}^{-1}\mathbf{C}_{,\boldsymbol{\beta}}\right]. \tag{17}$$

If we instead choose, $\mathbf{M} = \mathbf{F}^{-1}$, $\hat{p}_{\rm OQE}$ also has the desirable property that its window functions are Kronecker deltas so that no mixing between bandpowers occurs. However, fluctuations from the mean, described by the bandpower covariance matrix

$$\Sigma_{\alpha\beta} \equiv \langle \hat{p}_\alpha \hat{p}^*_\beta \rangle - \langle \hat{p}_\alpha \rangle \langle \hat{p}^*_\beta \rangle \tag{18}$$

are significantly larger and more correlated (Liu & Tegmark 2011).

Comparing equation (16) with equations (9) and (11), one can plainly see that the OQE is a result of choosing $\mathbf{R}^{\rm OQE} = \mathbf{C}^{-1}$ and $\mathbf{Q}_\alpha{}^{\rm OQE} = \mathbf{C}_{,\alpha}$.

3 DAYENU – A SIMPLE FOREGROUND FILTER

Unfortunately, many of the ingredients in equation (16) including $\mathbf{C}^{-1}$ weights, b_α, and $\mathbf{F}$, require perfect knowledge of $\mathbf{C}$ which includes thermal noise, the 21-cm signal, and instrumental effects such as antenna gains. Moreover, our understanding of the radio sky and radio interferometers is limited. We also do not really know what the correct $\mathbf{C}_{,\alpha}$ are either – the focus of paper II. In order to implement an OQE, several authors attempted to estimate $\mathbf{C}$ directly from the data. Dillon et al. (2015) obtained $\hat{\mathbf{C}}$, an estimate of $\mathbf{C}$ for the frequency–frequency covariance of 3D gridded visibilities by treating all other visibilities in an annulus of fixed u as independent samples of the same covariance, ignoring correlations in u. Ali et al. (2015) implemented a per-baseline OQE $\hat{\mathbf{C}}$ by computing the covariance between channels of an individual baseline over time. In that case, because $\hat{\mathbf{C}}$ is derived from the data itself, there exists significant risk of signal loss (Cheng et al. 2018). Loss issues led the PAPER team to seek simpler alternatives to $\mathbf{C}$ estimation. In their most recent analysis, PAPER implemented a per-baseline QE identical to a windowed Fourier transform with $\mathbf{R} = \mathbf{R}^{\rm BH}$, $\mathbf{M} = \mathbf{I} \equiv \mathbf{M}_{\rm ID}$, and $\mathbf{Q}_\alpha = \mathbf{Q}^{\alpha,{\rm DFT}}$ (Kolopanis et al. 2019).

Unfortunately, conservative taper-only filtering choices are of limited utility since they are unable to directly address the sidelobes from incomplete frequency sampling resulting from RFI flags. CLEANing provides a pre-processing option that can remove a significant fraction of this ringing but has the drawbacks that it is slow and the resulting statistics are difficult to propagate into a final estimate. Furthermore, under realistic flagging conditions, no implementation of 1D CLEAN has yet been shown to provide the level of foreground subtraction necessary for a robust 21-cm detection. Thus, relying on CLEAN is a significant risk. A second approach is to model the foreground covariance given our best understanding of the sky's statistics and our radio telescope. Works such as Shaw et al. (2014) and Trott et al. (2016) construct detailed models of diffuse and point-source foregrounds and incorporate information on the instrumental primary beam and antenna gains. Modelling approaches are a promising alternative to data-driven covariances that seemingly avoid the associated signal loss risks. However, it is not yet understood what amount of detailed modelling needs to be included in an inverse covariance filter for it to provide sufficient foreground suppression, especially when our knowledge of the instrument and radio sky are so limited. In this work, we explore a third option; modelling our covariance using as little knowledge of our telescope and foreground statistics as possible (DAYENU).

3.1 What makes a covariance model good enough?

Before we construct a simple covariance filter, we should get a sense of what the requirements on an inverse covariance filter are by writing down its action on a data vector.

If $\mathbf{Q}_\alpha$ performs an untapered Fourier transform, then any foregrounds that are left in our data at this point will be smeared by RFI gaps and the finite bandwidth. Thus, we want the ratio between foregrounds and signal in our inverse covariance-weighted data to be smaller than the level of sidelobes from finite bandwidth and RFI gaps.

To see what requirements this demand puts on our covariance model, we can decompose a hypothetical, non-singular covariance model $\hat{\mathbf{C}}$ into the sum of eigenvalue-weighted outer-products of its eigenvectors which we divide into a set that are dominated by signal $\{\mathbf{u}_s\}$ and a set that our dominated by foregrounds $\{\mathbf{u}_f\}$

$$\hat{\mathbf{C}} = \sum_s \lambda_s \mathbf{u}_s \mathbf{u}_s^\dagger + \sum_f \lambda_f \mathbf{u}_f \mathbf{u}_f^\dagger. \tag{19}$$

The action of $\hat{\mathbf{C}}^{-1}$ on a data vector $\mathbf{x}$ as

$$\begin{aligned}\mathbf{z} \equiv \hat{\mathbf{C}}^{-1}\mathbf{x} &= \sum_{\rm s} \frac{1}{\lambda_{\rm s}} \mathbf{u}_{\rm s} \left(\mathbf{u}_{\rm s}^\dagger \cdot \mathbf{x}\right) + \sum_{\rm f} \frac{1}{\lambda_{\rm f}} \mathbf{u}_{\rm f} \left(\mathbf{u}_{\rm f}^\dagger \cdot \mathbf{x}\right) \\ &= \sum_{\rm s} \frac{1}{\lambda_{\rm s}} \mathbf{u}_{\rm s} x_{\rm s} + \sum_{\rm f} \frac{1}{\lambda_{\rm f}} \mathbf{u}_{\rm f} x_{\rm f},\end{aligned} \tag{20}$$

where x_s are the coefficients of each signal-dominated mode in the data-vector and x_f are the coefficients of each foreground-dominated mode in the data. We see in equation (20) that all our inverse covariance weighting does is down-weights modes that we have identified as foregrounds in our covariance by λ_f and signal by λ_s. As long as λ_f is larger then λ_s by the dynamic range between the signal and the foregrounds, then $\mathbf{z}$ is dominated by signal. Note that it does not actually matter that we get the λ_f values right. They just have to be large enough to make the foreground terms much smaller then the signal terms. This is not typically difficult, especially since λ_f and λ_s square any estimate of the dynamic range between foregrounds and signal so even if an estimate of the dynamic range is low, it is made up for in the squaring.

We can go one step further and set $\lambda_s = 1$ so that our inverse covariance-weighted vector $\mathbf{z}$ includes signal modes with unity weight and foreground modes that are downweighted by $\lambda_f \gg 1$. As long as we come up with a model covariance whose foreground component is described a relatively small number of orthonormal modes and these modes span the actual foregrounds, the relative amplitudes of the foreground components in our covariance do not actually matter as long as they are large enough to suppress the foregrounds in the data below the signal. While this is a straightforward requirement, it means that regularization factors larger then the signal-foreground dynamic range will spoil foreground subtraction. For example, if $\hat{\mathbf{C}}$ includes the thermal noise component of a visibility after a short integration, as is the case in Dillon et al. (2015), Ali et al. (2015), and Trott et al. (2016), then it may actually prevent sufficient foreground subtraction for a 21-cm detection even though the covariance is technically more representative of the true data.

To summarize, we have shown that a $\hat{\mathbf{C}}$ is good enough for 21-cm power-spectrum estimation in the presence of missing data (RFI gaps and finite, untapered bandwidth) when it upweights all of the principal components of the foregrounds to larger then the dynamic range between foreground and signal modes in the data. The detailed amplitudes of each mode in the actual covariance does not matter as long as the dynamic range is large enough. Covariance models that include thermal noise for short integrations may not include sufficient dynamic range. We can avoid downweighting signal entirely by setting λ_s to unity in an estimated covariance by including only foreground modes with large λ_f added to an identity matrix.

In the remainder of this section, we will derive a simple covariance matrix that meets these requirements, motivated by the fact that foregrounds are overwhelmingly contained to large wavelength frequency Fourier modes over a finite range of delays. The covariance that we do derive will be diagonalized by DPSSs which are a set of vectors whose Fourier coefficients are maximally concentrated to within a finite delay-range. This basis is optimal in the sense that its vectors have maximal dot-products with foregrounds on large-frequency scales and minimal dot-products with the 21-cm signal at fine frequency scales and is an excellent choice for modelling and subtracting band-limited foregrounds in 21-cm experiments.

3.2 Defining DAYENU

As a first step towards understanding the necessary modelling fidelity required for effective foreground subtraction we attempt to write a model covariance that makes only the simplest assumptions about the foregrounds on an individual baseline. It has long been appreciated that if we could somehow take a continuous and infinite frequency Fourier transform of a visibility with an achromatic beam, that the power from spectrally flat foregrounds is completely contained to delays with amplitudes less then $\tau \leq \tau_H = b/c$, where c is the speed of light and b is the separation between the two antennas forming the visibility (Datta et al. 2010; Morales et al. 2012; Vedantham et al. 2012; Parsons et al. 2012b). Beam chromaticity and realistic spectral slope and curvature in the foregrounds modify this result but as long as these effects are relatively smooth (Ewall-Wice et al. 2016c; Thyagarajan et al. 2016; Patra et al. 2018), they still allow one to define some delay $\tau_w \gtrsim \tau_H$ below which foregrounds are much brighter than any 21-cm contribution and above which foregrounds are much smaller then both their $\tau = 0$ value and 21-cm fluctuations.

For a particular baseline, we make the simple assumption that the power in each delay is uncorrelated, an assumption that is true for point-source foregrounds but not strictly true for diffuse emission. This is because different delays map to different regions on the sky. Blake & Wall (2002) find source correlations fall below $\approx 10^{-3}$ on large scales greater then 1°, thus the different delays for different regions are approximately uncorrelated. Since diffuse emission in different regions of the sky is correlated, diffuse emission in different delays is correlated. In order for delays to be uncorrelated, we must also impose an assumption that the statistics in frequency space are stationary (frequency independent).

When $\tau \leq \tau_w$ (foreground region), we assume that the variance of each delay is the inverse of a small number ϵ. For $\tau \geq \tau_w$, we set the variance equal to the channel-width $\Delta\nu$

$$\widetilde{C}^{\urcorner}(\tau, \tau') = \begin{cases} \epsilon^{-1} \frac{1}{2\tau_w} \delta^D(\tau - \tau') & |\tau| \leq \tau_w \\ \Delta\nu\, \delta^D(\tau - \tau') & |\tau| > \tau_w. \end{cases} \tag{21}$$

Here, $\Delta\nu$ is the width of each frequency channel and not necessarily the spacing between different channels. The first piece of equation (21) represents foregrounds in delay-space while the second piece represents thermal noise.

Suppose we have measurements at N_d different arbitrary frequencies. The covariance matrix for these discrete measurements can be obtained by integrating the continuous delay covariance

$$\begin{aligned} C^{\urcorner}_{mn} &= \int \mathrm{d}\tau \mathrm{d}\tau' \mathrm{e}^{-2\pi i(\tau\nu_j - \tau'\nu_k)} \widetilde{C}^{\urcorner}(\tau, \tau') \\ &= \epsilon^{-1} \mathrm{Sinc}\left[2\pi\tau_w(\nu_m - \nu_n)\right] + \Delta\nu\delta^D(\nu_m - \nu_n) \\ &= \epsilon^{-1} \mathrm{Sinc}\left[2\pi\tau_w(\nu_m - \nu_n)\right] + \delta^k_{mn}, \end{aligned} \tag{22}$$

where $\mathrm{Sinc}[x] \equiv \sin x/x$. In the last line of equation (22), we substitute the Dirac delta-function for a Kronecker delta,[3] $\Delta\nu\delta_D \to \delta^k$. An astute reader might note that we could have just as easily have constructed $\widetilde{\mathbf{C}}^{\urcorner}$ as being diagonal in *discrete* delay space instead of continuous delay space and constructed $\mathbf{C}^{\urcorner}$ by taking the 2D DFT of $\widetilde{\mathbf{C}}^{\urcorner}$ instead of performing the integrals in equation (22). We will justify our choice of a continuous definition in Section 3.6 but for now we emphasize that defining $\widetilde{\mathbf{C}}^{\urcorner}$ in continuous delay-space is essential to its efficacy.

In equation (22), we assumed that foregrounds uniformly occupy a finite range of delays between $-\tau_w$ and τ_w. More generally, we can model foregrounds occupying any number of rectangular delay regions (indexed by ℓ) with half widths of τ^ℓ_w centred at τ^ℓ_c and uniform amplitude ϵ_ℓ.

$$C^{\urcorner}_{mn} = \delta^k_{mn} + \left[\mathbf{C}^{\urcorner}_{\mathrm{FG}}\right]_{mn}, \tag{23}$$

where

$$\left[\mathbf{C}^{\urcorner}_{\mathrm{FG}}\right]_{mn} = \sum_\ell \frac{1}{\epsilon_\ell} \mathrm{e}^{-2\pi i \tau^\ell_c(\nu_m - \nu_n)} \mathrm{Sinc}\left[2\pi\tau^\ell_w(\nu_m - \nu_n)\right]. \tag{24}$$

[3]This standard normalization for replacing the Dirac delta with the Kronecker delta ensures that $1 = \int \mathrm{d}\nu\delta_D = \Delta\nu \sum_n \delta^K_{mn}/\Delta\nu$.

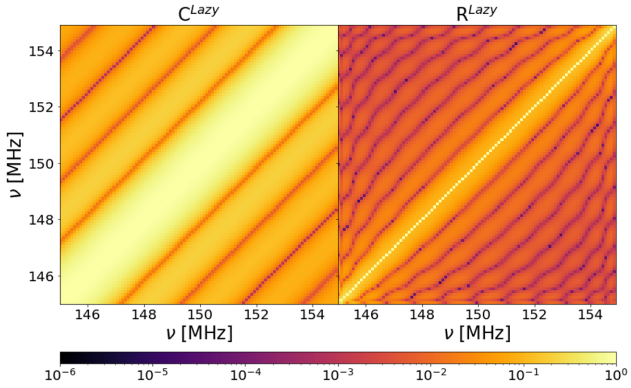

Figure 1. Left: An example of $\mathbf{C}^{\urcorner}$ for 100 channels, $\Delta\nu = 100\,\mathrm{kHz}$, $\epsilon = 10^{-9}$, and $\tau_w = 250\,\mathrm{ns}$. $\mathbf{C}^{\urcorner}$ is a covariance that is diagonal in the continuous Fourier basis and as a result is Toeplitz. Right: To obtain a filter matrix, we take the inverse of $\mathbf{C}^{\urcorner}$ and obtain $\mathbf{R}^{\urcorner}$. While this inverse is translation invariant in the limit of infinite frequency resolution, it is not for discrete channels.

A covariance with multiple delay regions, such as the one in equation (24) can be useful for filtering data with super-horizon artefacts including cable reflections (Dillon et al. 2015; Beardsley et al. 2016; Ewall-Wice et al. 2016b).

We define our lazy DAYENU filter to be the inverse of $\mathbf{C}^{\urcorner}$

$$\mathbf{R}^{\urcorner} = [\mathbf{C}^{\urcorner}]^{-1}. \tag{25}$$

While $\mathbf{C}^{\urcorner}$ is Toeplitz, the actual weighting that we apply to visibility data, $\mathbf{R}^{\urcorner}$ is not (Fig. 1).

3.3 Without RFI flags, $\mathbf{C}^{\urcorner}$ is diagonalized by discrete prolate spheroidal sequences

The Sinc foreground component to the covariance in equation (22) is diagonalized by a heavily studied set of orthonormal vectors known as discrete prolate spheroidal sequences (DPSSs, Slepian 1978).

Letting $\mathcal{W} = \tau_w \Delta\nu$, Slepian (1978) define a DPSS $\boldsymbol{u}^{(\alpha)}(N_\mathrm{d}, \mathcal{W})$ to be one of the countable orthonormal set of vectors solving the eigenvalue problem

$$\sum_{n=0}^{N_\mathrm{d}-1} L_{mn}(N_\mathrm{d}, \mathcal{W}) u_n^{(\alpha)}(N_\mathrm{d}, \mathcal{W}) = \lambda_\alpha(N_\mathrm{d}, \mathcal{W}) u_m^{(\alpha)}(N_\mathrm{d}, \mathcal{W}) \tag{26}$$

where

$$L_{mn}(N_\mathrm{d}, \mathcal{W}) = \frac{\sin 2\pi\mathcal{W}(m-n)}{\pi(m-n)}. \tag{27}$$

Since $\mathbf{L} = 2\mathcal{W}\mathbf{C}^{\urcorner}_{\mathrm{FG}}$, the DPSSs also diagonalize $\mathbf{C}^{\urcorner}_{\mathrm{FG}}$. Because $\mathbf{C}^{\urcorner}$ is the sum of $\mathbf{C}^{\urcorner}_{\mathrm{FG}}$ and an identity term, DPSSs are also the eigenvectors of $\mathbf{C}^{\urcorner}$ as we show numerically in Fig. 2. Let $\{h_n\}_{N_\mathrm{d}}$ be the set of all complex sequences of length N_d. Slepian (1978) shows that $\boldsymbol{u}^{(0)}(N_\mathrm{d}, \mathcal{W})$ the DPSS with the largest eigenvalue λ_0 is the unit-norm N_d sequence that maximizes the quantity

$$\mu \equiv \frac{\int_{-\mathcal{W}}^{\mathcal{W}} |H(f)|^2 \mathrm{d}f}{\int_{-1}^{1} |H(f)|^2 \mathrm{d}f}, \tag{28}$$

where $H(f)$ is the DFT of h_n centred at $n = (N_\mathrm{d} - 1)/2$

$$H(f) = \mathrm{e}^{-i\pi f(N_\mathrm{d}-1)} \sum_{n=0}^{N_\mathrm{d}-1} \mathrm{e}^{-2\pi i n f} h_n. \tag{29}$$

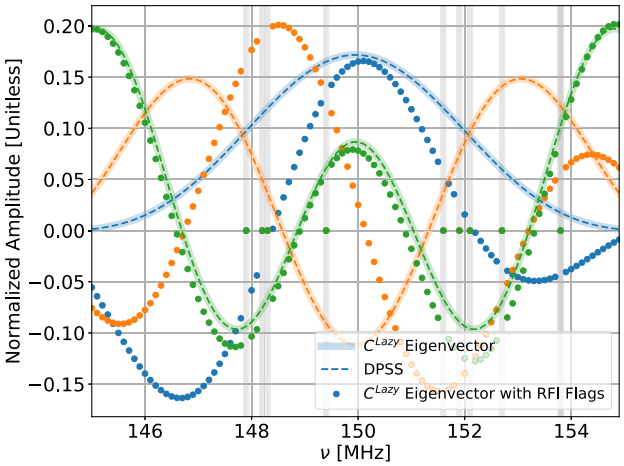

Figure 2. The eigenvectors of the $\mathbf{C}^{\urcorner}$ in Fig. 1 with $N_\mathrm{d} = 100$, $B = 10\,\mathrm{MHz}$, $\tau_w = 150\,\mathrm{ns}$, and $\epsilon = 10^{-9}$ for the zeroth (blue), second (orange), and fourth (green) largest eigenvalues (wide light lines). We compare these eigenvectors to the zeroth (blue), second (orange), and fourth (green) DPSSs of length $N_\mathrm{d} = 100$, $\tau_w = 150\,\mathrm{ns}$, over a frequency bandwidth of $B = 10\,\mathrm{MHz}$ (dashed lines). With no flags present $\mathbf{C}^{\urcorner}$ is diagonalized by DPSSs. We next set 10 random rows and columns of $\mathbf{C}^{\urcorner}$ equal to zero to simulate RFI flags. The resulting eigenvectors (dotted lines) do not correspond to DPSSs.

They also show that $\boldsymbol{u}^{(1)}(N_\mathrm{d}, \mathcal{W})$ is the vector that simultaneously maximizes μ, has unity norm, and is orthogonal to $\boldsymbol{u}^{(0)}(N_\mathrm{d}, \mathcal{W})$. More generally, $\boldsymbol{u}^{(\alpha)}(N_\mathrm{d}, \mathcal{W})$ is the vector that simultaneously maximizes μ, has unity norm, and is orthogonal to the vectors in the set $\{\boldsymbol{u}^{(\alpha')}(N_\mathrm{d}, \mathcal{W}) : \alpha' < \alpha\}$.

It follows that DPSSs have the ideal property of maximally concentrating power into a rectangular region of Fourier space with half-bandwidth τ_w. The DPSS with the largest eigenvalue is the unity norm N_d length sequence that concentrates maximal power (as quantified by μ) within τ_w. The DPSS with the second largest eigenvalue is the unity norm N_d-length sequence that maximally concentrates power within τ_w and is orthogonal to the DPSS with the largest eigenvalue. Ordering DPSSs by their eigenvalues (largest to smallest), the αth DPSS for N_d and τ_w is the length N_d unity-norm sequence that maximally concentrates power within τ_w and is orthogonal to all $\alpha' < \alpha$ DPSSs. Thus, our foreground covariance is diagonalized by the basis that most efficiently concentrates power within $\tau < \tau_w$. In the absence of channel flags, DPSS vectors are the eigenbasis of $\mathbf{C}^{\urcorner}$. As we discussed in Section 3.1 though this covariance may not include the detailed information on the true values of λ_f for each foreground mode on a particular baseline, as long as ϵ^{-1} is large enough, it will remove the foregrounds to a small enough level that we can measure the 21-cm signal in the presence of flagging sidelobes.

Slepian (1978) also shows that the first $\approx 2N_\mathrm{d}\mathcal{W}$ eigenvalues of $\mathbf{L}$, $\lambda_\alpha(N_\mathrm{d}, \mathcal{W})$, are close to unity after which they rapidly drop to zero. When N_d is small, the number of non-zero eigenvalues tends to exceed this number but it becomes increasingly accurate as N_d increases. Fitting and characterizing foregrounds with DPSS vectors therefor requires $\approx 2B\tau_w$ components.

Under the realistic circumstance that there is missing data (e.g. RFI gaps), the eigenvectors are not equal to DPSSs. In Fig. 2, we compare the zeroth, second, and fourth numerically determined eigenvectors (ordered by decreasing eigenvalue) of $\mathbf{C}^{\urcorner}$ in Fig. 1 to DPSSs with length N_d, frequency bandwidth $B = 10\,\mathrm{MHz}$, and delay-space width of $\tau_w = 150\,\mathrm{ns}$. To within numerical precision, the

DPSSs are identical to numerically computed eigenvectors of $\mathbf{C}^{\urcorner}$. We flag 10 random channels in $\mathbf{C}^{\urcorner}$ by setting the corresponding rows and columns to zero and show the resulting eigenvectors with the zeroth, second, and fourth largest eigenvalues. The eigenvectors of $\mathbf{C}^{\urcorner}$ with flagged channels are not merely DPSSs with flagged elements equal to zero. Hence, when we have missing data (RFI gaps), we must set the corresponding rows and columns of $\mathbf{C}^{\urcorner}$ to zero and set $\mathbf{R}^{\urcorner}$ equal to the psuedo-inverse of this flagged covariance.

As stated in Section 3.1, the effective action of $\mathbf{R}^{\urcorner}$ is to transform our data into a basis close to DPSSs where $\mathbf{C}^{\urcorner}$ is diagonal, divide the data by the eigenvalues of $\mathbf{C}^{\urcorner}$ in the $\mathbf{C}^{\urcorner}$ eigenbasis, and then transform back. The degree to which foreground removal and signal preservation are successful depends on how well isolated foreground and signal components are in the $\mathbf{C}^{\urcorner}$ eigenbasis and whether we have included sufficient dynamic range in the ϵ^{-1} parameter of $\mathbf{R}^{\urcorner}$.

3.4 A simple example

As a first test, we apply it to a realization of a simplistic model autocorrelation for an isotropic sky with temperature $T_{\rm sky} = 60\,{\rm K}\,(\lambda/1\,{\rm m})^{2.55}$, a chromatic Airy beam from a 14-m diameter aperture, a receiver temperature of 100 K, and 200, evenly spaced frequency channels, of width $\Delta\nu = 100$ kHz between 140 and 160 MHz. To simulate RFI flags, we randomly set the power levels in 20 channels to zero. To simulate thermal noise, we assume an integration time of $t_{\rm int} = 100$ h, similar to what is necessary for a robust 21-cm detection, and set the standard deviation of each channel equal to $A/\sqrt{\Delta\nu t_{\rm int}}$ where A is the autocorrelation amplitude (Thompson, Moran & Swenson 2017). In Fig. 3, we show the impact of applying $[\mathbf{C}^{\urcorner}]^{-1}$ to a single realization of the autocorrelation with $\epsilon = 10^{-10}$ and $\tau_w = 50$ ns. After applying our filter, the foregrounds are suppressed by six orders of magnitude and the remaining residual (orange line) is very close to the original noise (green line). Taking the difference between the injected noise and residuals (dotted grey) we see that in the frequency domain, the filter residuals agree with the injected noise at the ≈10 per cent level.

In the bottom panel of Fig. 3, we inspect our simulation in the delay domain. In the absence of flags, we can use a 7-term Blackman–Harris[4] taper-filtered Fourier transform to suppress the impact of a finite sampling bandwidth beyond ≈250 ns (solid grey line). When we set channels containing RFI to zero, these sharp edges spread foregrounds across all DFT modes (black dashed line). We compare the Blackman–Harris Fourier transform of residuals after applying $\mathbf{R}^{\urcorner}$ and the injected noise in delay space. The majority of the ≈10 per cent disagreement observed in frequency space is contained within 250 ns of the edge of our filter (shaded grey region).

Beyond 250 ns the injected noise and $\mathbf{R}^{\urcorner}$ residuals agree at the ≈10 per cent level. At $\tau \gtrsim 250$ ns, the leaked foregrounds are subtracted to the level of 10^{-8}, even with flagging. This is much better than what can be accomplished by an apodized DFT with no flagging. Since apodization functions go to zero at the band edges, they also attenuate the signal. While we applied an apodization before DFTing $\mathbf{R}^{\urcorner}\mathbf{x}$ to obtain a more direct comparison with the unflagged

[4]The 7-term Blackman-Harris (see, for example Solomon 1993) includes additional sinusoidal terms beyond the standard 4-term Blackman-Harris found in standard libraries such as `scipy.signal` (Virtanen et al. 2020). While the additional terms increase the width of the central lobe, they substantially lower sidelobes compared to the typical 4-term implementation. We use a 7-term Blackman-Harris taper for all analysis in this paper and refer to it hereon out as simply 'Blackman-Harris'.

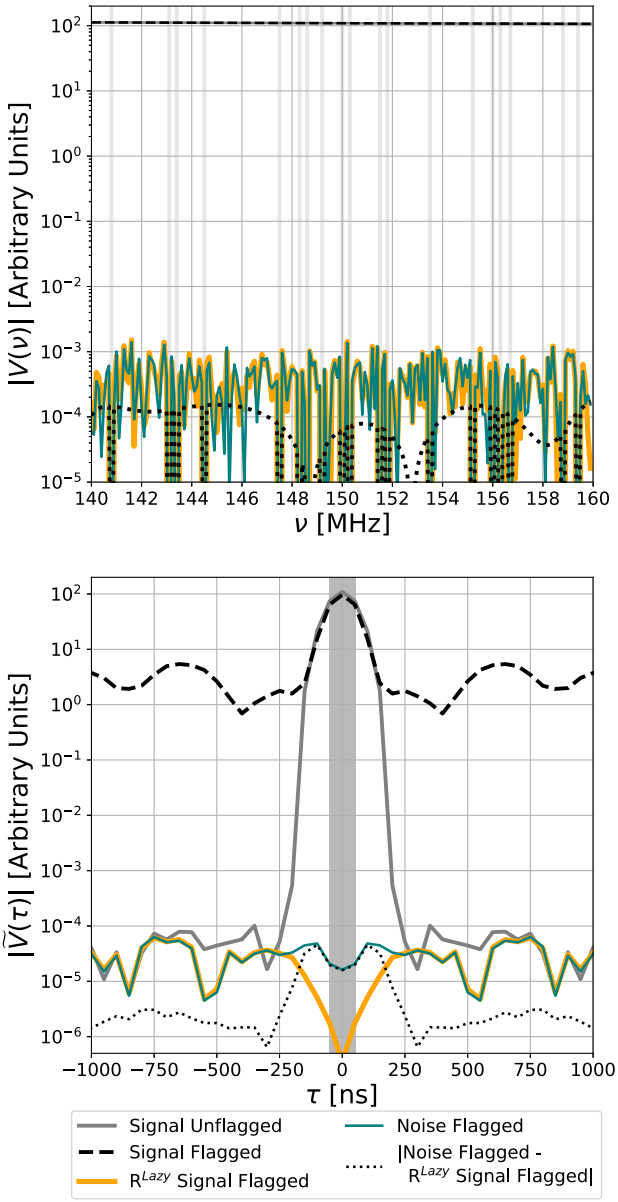

Figure 3. Top: A simulated signal with 200 channels (noise plus foregrounds) at a single LST drawn from a Gamma distribution with variance consistent with 100 h of integration, similar to what is necessary for a 21-cm detection, with (dashed black line) and without (solid grey line) 20 random flags. Flagged channels are shown with vertical grey lines and the corresponding rows and columns in $\mathbf{C}^{\urcorner}$ are set to zero before calculating the psuedo-inverse for $\mathbf{R}^{\urcorner}$. Channel–channel fluctuations (thermal noise) are at the $\sim 10^{-5}$ level (orange line). Residuals after applying $\mathbf{R}^{\urcorner}$ with $\tau_w = 50$ ns, $\epsilon = 10^{-9}$ to the flagged signal results in the teal curve. The difference between $\mathbf{R}^{\urcorner}$ residuals and the injected noise at the 10 per cent level (dotted black line). Bottom: the same as the top but in the DFT domain (with Blackman–Harris windowing). The filter residual agrees very well with the noise (compare teal and orange in both plots) except for within 100–200 ns of the attenuation region (shaded grey rectangle in bottom panel) where some foreground residual is still present. `DAYENU` does not have to down-weight power near the band edges, leading to similar levels of foreground residual across the entire band (dotted black line). Outside of ∼200 ns, the noise is preserved by the filter at the level of a few per cent (compare black dotted and orange lines).

model in which no foregrounds were filtered, we technically did not have. Thus, applying $\mathbf{R}^{\urcorner}$ allows one to circumvent the band-edge signal attenuation that comes with apodization.

In this simplified example, $\mathbf{R}^{\urcorner}$ is highly effective at suppressing foregrounds. However, our simulation made a number of unrealistic assumptions. We assumed an isotropic sky with identical spectral indices. In addition, we assumed that the only chromaticity in our antenna response was sourced by its airy function beam pattern. Ultimately, $\mathbf{R}^{\urcorner}$ and any other inverse covariance filter schemes will only be effective if the foregrounds as viewed by the instrument are spanned by the model covariance's foreground eigenmodes and the model covariance has enough dynamic range to suppress the foreground modes in the data to a level where their flagging sidelobes do not mask the 21-cm signal power spectrum. For $\mathbf{R}^{\urcorner}$, this means that it will prevent foreground bleed by the DFT and missing data as long as ϵ is large enough and τ_w extends beyond the delays where the foregrounds convolved with the instrument exceed the 21-cm signal level. From a practical standpoint, this means that $\mathbf{R}^{\urcorner}$ cannot help us detect 21-cm fluctuations if internal and external antenna reflections as observed for example by Beardsley et al. (2016), Ewall-Wice et al. (2016a), and Kern et al. (2019) extend into the delays where interferometers derive most of their sensitivity. On the other hand, if the signal chain chromaticity is contained within some upper τ_w; a design requirement for the Hydrogen Epoch of Reionization Array (HERA; DeBoer et al. 2017), then all an analyst needs to do in order to filter foregrounds from their data is to choose a large ϵ^{-1} and set an appropriate τ_w in $\mathbf{R}^{\urcorner}$ that extends to the horizon delay $\tau_{\rm H}$ plus the intrinsic chromaticity of the antenna. Considering HERA as an example; the HERA antenna's chromaticity leaks power above ≈ -50 dB at 250 ns (Ewall-Wice et al. 2016c; Thyagarajan et al. 2016; Patra et al. 2018). For HERA, we therefore recommend a τ_w equal to the wedge plus roughly 250 ns.

3.5 Filtering efficacy and signal attenuation

To be an effective foreground filter, $\mathbf{R}^{\urcorner}$ should attenuate foregrounds while leaving as much of the 21-cm signal as untouched as possible. If 21 cm is also attenuated and we do not account for this attenuation in the normalization step we can end up with an unaccounted bias in our measurement: signal loss. Signal loss is not necessarily a bad thing and is in fact desirable if it suppresses foregrounds on otherwise contaminated 21-cm modes (we would not want our normalization to restore this). In paper II, we will explore when and how good signal loss occurs. In this paper, we focus on the attenuation properties of our simple filter DAYENU with the conservative assumption that we use $\mathbf{M}_{\rm ID}$ so no correction is made at the normalization step. Under these conditions we treat signal attenuation as significant if its power-spectrum signature exceeds sample variance errors which dominate the most sensitive regions of k-space in upcoming experiments. Lanman & Pober (2019) find that sample variance errors for per-baseline power spectra are of the order of 20 per cent which places a 10 per cent constraint on attenuation in the visibility domain. Spherically averaged power spectra are expected to be far more sensitive, with $\sim$2 per cent sample-variance errors. This places a constraint of 1 per cent on visibility attenuation.

We investigate the degree that DAYENU can suppress modes with different τ by studying the amplitudes of $\mathbf{z}^\tau = \mathbf{R}^{\urcorner}\mathbf{x}^\tau$ where $\mathbf{x}^\tau$ is a complex sinusoid with delay τ and amplitude equal to unity sampled every 100 kHz. In Fig. 4, we plot the RMS of $\mathbf{z}^\tau$, $\sqrt{N_{\rm d}^{-1}\sum_m |z^\tau_m|^2}$ versus τ for two bandwidths; 10 and 100 MHz, $\epsilon = 10^{-9}$, and two filter widths; $\tau_w = 150$ and $\tau_w = 500$ ns.

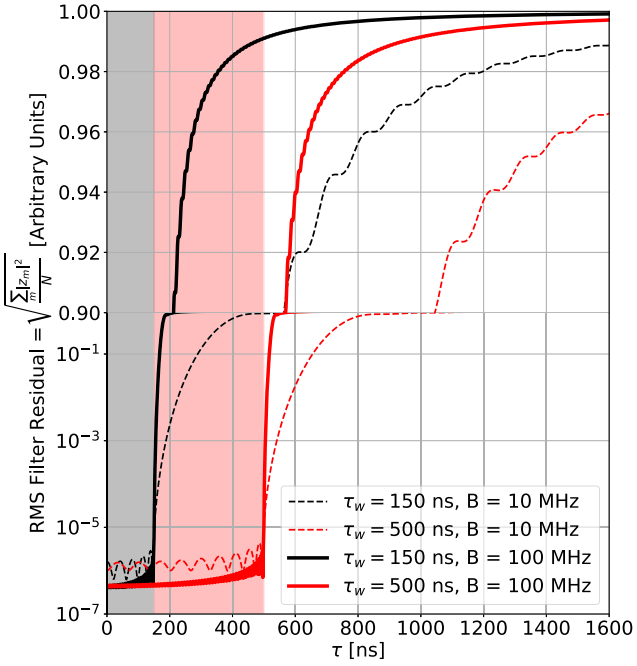

Figure 4. The RMS of residual after applying a $\mathbf{R}^{\urcorner}$ with $\epsilon = 10^{-9}$; $\tau_w = 150$ ns (black lines), and $\tau_w = 500$ ns (red lines); and bandwidths of 10 (dashed lines) and 100 MHz (solid lines). Note that the bottom panel has a logarithmic y-scale and the top panel has a linear y-scale. Shaded regions indicate the τ_w half widths of each filter. Tones within the attenuation region are suppressed between 10^{-7} and 10^{-6}, more than enough for robust 21-cm studies. Greater filter bandwidth allows for enhanced overall suppression and reduces attenuation outside of the attenuation region. Attenuation above 10 per cent is required to bring biases below the level of expected sample variance in per-baseline power spectrum estimates. This occurs for $\tau \gtrsim 300$ ns beyond the filter edge if a filtering bandwidth of 10 MHz is used and only 50 ns beyond the filter edge if a bandwidth of 100 MHz is employed. Spherical power spectrum estimates will bring variance errors down to 2 per cent in the power spectrum which translates to a 1 per cent attenuation requirement in visibility space. Filtering over 100 MHz brings attenuation below 1 per cent for $\tau \gtrsim 300$ ns beyond the filter edge with 100 MHz of filtering bandwidth. In principal, attenuation can be corrected for at the power-spectrum normalization step so these requirements only strictly apply to power spectrum estimates with identity normalization.

Within the attenuation region, we see that input tones are suppressed by a factor of 10^{-7}–10^{-6}, depending on the bandwidth with larger bandwidths achieving more effective suppression. When 10 MHz of bandwidth is used, $\gtrsim$10 per cent signal attenuation occurs within roughly 300 ns of the filter edge. Performance improves dramatically if a filtering bandwidth of 100 MHz is used instead. For 100 MHz filtering, $\lesssim$10 per cent attenuation occurs beyond 50 ns of the filter edge and $\lesssim$1 per cent attenuation is reached by 300 ns beyond the filter edge. Thus, if we conservatively choose to normalize with $\mathbf{M}_{\rm ID}$ then attenuation beyond 300 ns will be smaller then the expected sample variance errors in upcoming experiments. $\mathbf{M}_{\rm ID}$ is a conservative choice, however, and we can do better if we choose normalizations that undo these attenuations which we explore in paper II.

We also inspect how the amplitude of $\mathbf{z}^\tau$ depends on ϵ in Fig. 5. We note that the overall level of suppression is consistent (within a few dB) whether we filter across 100 or 10 MHz. We compute the average level of suppression of tones over a range of τ_w and bandwidths as a function of ϵ in Fig. 6. For a fixed ϵ, the amplitudes of residuals

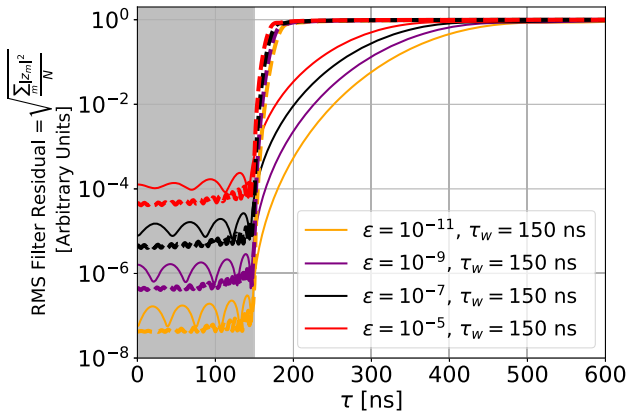

Figure 5. The RMS residual after applying $\mathbf{R}^{\urcorner}$ across 100 (solid lines) and 10 MHz (dashed lines) for different values of ϵ. The level of suppression within the filter region is roughly consistent within 0.25 dex for fixed ϵ and different bandwidths.

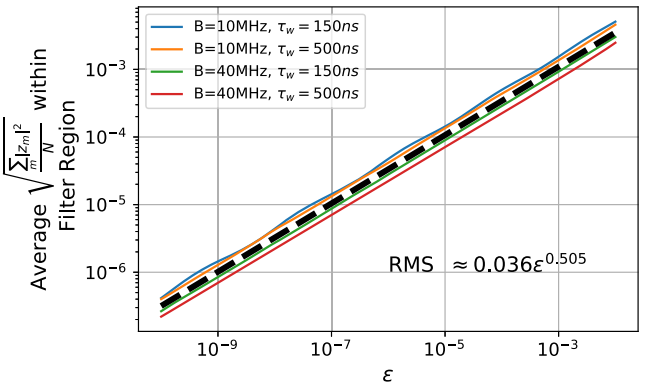

Figure 6. The average suppression of tones within the filter region induced by $\mathbf{R}^{\urcorner}$ for several different filtering bandwidths and τ_w values (coloured lines) as well as a power-law fit to their average (black dashed line). For fixed ϵ, the residual amplitudes inside of the filter region agree within a few dB over a wide range of τ_w and bandwidths. The RMS residual amplitude goes roughly as the square root of ϵ (dashed lines). It follows that $\epsilon \lesssim 10^{-8}$ should be used to reduce foreground residuals by a factor of $\approx 10^{-5}$, comfortably below the 21-cm signal.

within the filtering region agree within 0.25 dex over a wide range of τ_w and bandwidths. The RMS suppression of Fourier tones within the filtering region follows a power law which we fit to be RMS $\approx 0.1\epsilon^{0.5}$. It follows that to suppression 21 cm foregrounds that are $\approx 10^4$ times larger then cosmological fluctuations, we should apply filters with $\epsilon \lesssim 10^{-8}$. Since the foregrounds in the EoR window will be suppressed by flagging sidelobes, it is possible that one could get away with ϵ one-to-two orders of magnitude larger depending on the severity of flagging.

3.6 DAYENU and the DFT basis

To derive $\mathbf{C}^{\urcorner}$ (equation 22), we wrote down discrete elements of our frequency covariance matrix by taking the continuous Fourier transform of a covariance that was diagonal in continuous delay space. On the other hand, many power-spectrum estimators (e.g. Parsons et al. 2012b; Dillon et al. 2013; Trott et al. 2016; Barry et al. 2019) estimate band-powers in DFT space. This difference in approach immediately raises the question, why not derive $\mathbf{R}$ from a covariance matrix that is diagonal in DFT space rather than the continuous space that we chose? After all, if we could just write down $\mathbf{R}$ as diagonal in DFT space, could we just divide the DFT of our data set by the diagonal DFT of $\mathbf{R}$, $\widetilde{\mathbf{R}}$, and save computational steps? The short answer is that an $\mathbf{R}$ that is diagonal in DFT space only includes information on foreground modes with delays equal to m/B, $m \in \{-N_{\rm d}/2, \ldots N_{\rm d}/2 - 1\}$ and as a result is incapable of properly suppressing foregrounds at intermediate delays. In order to see this effect, we write $\mathbf{C}^{\rm DFT}$ as the DFT of a covariance that is diagonal in DFT space

$$\widetilde{C}^{\rm DFT}_{rs} = \begin{cases} \epsilon^{-1}\frac{1}{2\tau_w B}\delta^k_{rs} & \left|\frac{r}{B}\right| \leq \tau_w \\ \delta^k_{rs} & \left|\frac{r}{B}\right| > \tau_w \end{cases}. \tag{30}$$

We then transform $\widetilde{\mathbf{C}}^{\rm DFT}$ into discrete frequency space by performing a 2D DFT

$$\begin{aligned} C^{\rm DFT}_{mn} &= \delta^k_{mn} + \frac{\epsilon^{-1}}{2\tau_w B}\sum_{\substack{|r|\leq\tau_w B \\ |s|\leq\tau_w B}} {\rm e}^{-2\pi i(rm-sn)/N_{\rm d}}\delta^k_{rs} \\ &= \delta^k_{mn} + \frac{\epsilon^{-1}}{2\tau_w B}\sum_{|r|\leq\tau_w B} {\rm e}^{-2\pi ir(m-n)/N_{\rm d}} \\ &= \delta^k_{mn} + \epsilon^{-1}\sum_{s=-\infty}^{\infty} {\rm Sinc}\left[2\pi\tau_w\left(B\frac{m-n}{N_{\rm d}} - sB\right)\right], \end{aligned} \tag{31}$$

where we used the Poisson summation formula (e.g. Epstein 2007) to go from the second and third lines in equation (31). We see that the foreground component of $\mathbf{C}^{\rm DFT}$ is essentially an infinite sum of copies of the foreground component of $\mathbf{C}^{\urcorner}$ translated along the diagonal by integer multiples of B. This can also be seen by visual inspection in Fig. 7 where we plot $\mathbf{C}^{\urcorner}$ next to $\mathbf{C}^{\rm DFT}$. The wrap-around arises from the fact that our covariance elements are exclusively comprised of tones that are periodic over the interval B.

By definition, $\mathbf{C}^{\rm DFT}$ is diagonalized by the DFT. Thus, when we weight by its inverse, it will only down-weight modes with $\tau = mB^{-1} \leq \tau_w$; harmonic or on-grid DFT tones. Visibilities include a continuum of delays and only a fraction of their power is accounted for by harmonic tones within the wedge. Thus, $\mathbf{R}^{\rm DFT} \equiv [\mathbf{C}^{\rm DFT}]^{-1}$ is incapable of removing the bulk of foreground power, especially power in the sinc-sidelobes of the aharmonic tones. These sidelobes remain at high delays and prohibit a 21-cm measurement.

Fig. 8 illustrates the limitations of $\mathbf{C}^{\rm DFT}$, where we show the same quantities as in Fig. 4 but now include the performance of $\mathbf{R}^{\rm DFT}$. We study the impact of progressively adding in-between-modes back into $\mathbf{C}^{\rm DFT}$ by increasing the wrap-around interval in equation (31). For example, increasing the wrap-around from B to $2B$, adds additional modes that are periodic over a bandwidth of $2B$ but are not periodic over B. The orange lines in Fig. 8 show the residual amplitudes leftover after applying $\mathbf{R}^{\rm DFT}$ to complex sinusoides with various delays, τ. Unlike $\mathbf{R}^{\urcorner}$, gaps are present, $\mathbf{R}^{\rm DFT}$'s filter coverage and truely effective filtering only occurs at $\tau = m/B, m \in \mathbb{Z}$. Between B^{-1} harmonics, filtering only decreases the foreground amplitude by a factor of $\sim 10^{-1}$.

As we increase period of the wrap-around in equation (31), the harmonic filter tones move closer together and eventually merge. Because larger bandwidths have greater Fourier resolution, increasing the DFT wrap-around to 2B over 100 MHz actually attains similar performance for the completely continuous case though DAYENU still subtracts foregrounds to roughly $\approx 10^{-2}\times$ the level of DFT modes at the filter edge. This indicates that if we did want to use DFT modes to model our foregrounds and subract them, we need of the order of $\gtrsim 2\times$ as many modes. Since $\mathbf{C}^{\rm DFT}$ converges to DAYENU as the wrap

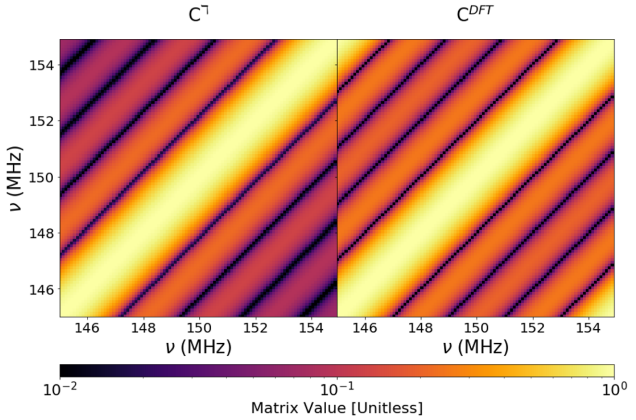

Figure 7. Left: $\mathbf{C}^{\urcorner}$ with $\tau_w = 250$ ns and $\epsilon = 10^{-9}$ where each element is obtained using a continuous Fourier transform (equation 22). Right: $\mathbf{C}^{\rm DFT}$, the 2D DFT of which is diagonal. The two matrices differ through the presence of wrap-around, $\mathbf{C}^{\rm DFT}$ is equal to an infinite sum of copies of $\mathbf{C}^{\urcorner}$ translated by integer intervals of B along the diagonal (equation 31). The real-life absence of correlations between opposite band-edges in our foregrounds, which is demanded by DFT modes, is what causes $\mathbf{C}^{\rm DFT}$ to perform poorly relative to $\mathbf{C}^{\urcorner}$.

interval approaches $\gtrsim 2B$, roughly $\gtrsim 4\tau_w B$ DFT modes are necessary to model foregrounds at a level similar to $\approx 2\tau_w B$ DPSS vectors. As we mentioned in Section 3.3, for large $N_{\rm d}$, the number of DPSS modes with non-zero eigenvalues in $\mathbf{C}^{\urcorner}$ is approximately $2B\tau_w$.

If the DPSS modes are precomputed and the number of DPSS modes being fit is much less then the number of frequency channels, then finding the fit coefficients for a single flagging pattern and set of fitted modes is dominated by calculating $\mathbf{A}^{\dagger}\mathbf{w}\mathbf{A}$ where $\mathbf{A}$ is the $N_{\rm d} \times N_{\rm mode}$ design matrix where each row is one of the $N_{\rm mode}$ DPSS vectors that we are fitting. This matrix multiplication requires $\sim\mathcal{O}(N_{\rm d}N_{\rm mode}^2)$ operations. Since typically twice as many DFT modes are required then DPSS modes, DPSS fitting with pre-computed modes reduces computational operations by a factor of four.

In summary, filtering with a covariance that is diagonal in the discrete Fourier basis will perform very poorly in foreground subtraction because it only contains the subset of foreground modes that are harmonics of B^{-1}. In defining $\mathbf{C}^{\urcorner}$, we instead allow foregrounds to include any continuous delay within the wedge and use numerical matrix inversion determine and downweight a discrete set of principal components.

3.7 Pre-truncation filtering

It is clear from Fig. 4 that the larger the bandwidth we filter over, the smaller the unwanted signal attenuation outside of τ_w. This motivates the use of ∼100-MHz bandwidths for filtering. The power spectrum is usually approximated over bandwidths of $\lesssim$10 MHz in order to ensure roughly stationary statistics for the evolving 21-cm signal.

These two ends can simultaneously be achieved by applying $\mathbf{R}^{\urcorner}$ over a ≈100-MHz band, truncating, and then estimating the power spectrum from a DFT over a smaller sub-band. Under this scheme, $\mathbf{R}^{\urcorner}$ is a non-square $N_{\rm d} \times N_{\rm d}^{\rm F}$ matrix, where $N_{\rm d}^{\rm F}$ is the number of channels to be filtered over and $N_{\rm d}^{\rm F} \geq N_{\rm d}$. To obtain a truncated $\mathbf{R}^{\urcorner}$, all we have to do is zero out the rows of $\mathbf{R}^{\urcorner}$ corresponding to channels that we do not want to include in the application of $\mathbf{Q}_{\alpha}$.

Fig. 9 examines signal attenuation as a function τ over 10 different 10 MHz sub-bands where truncation to 10 MHz is performed after the application of $\mathbf{R}^{\urcorner}$. In each sub-band, signal attenuation is dramatically reduced compared to filtering over the 10-MHz band alone. With the exception of the edge bands (100–110 and 190–200 MHz), $\lesssim$1 per cent signal attenuation is achieved by 250 ns

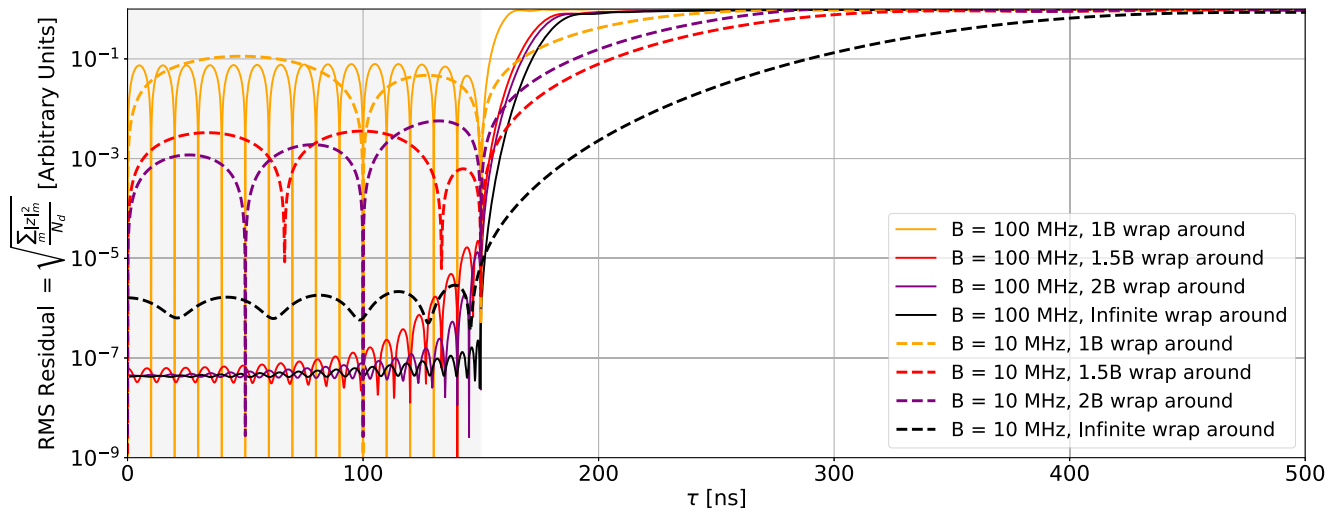

Figure 8. The RMS residual $\mathbf{R}^{\urcorner}$ applied of tones with delay τ. We filter $\tau \lesssim \tau_w \approx 125$ ns over 10 (dashed lines) and 100 MHz (solid lines) with the covariance matrix periodicity (the coefficient next to 'm' in equation 31) set to be $1B$ (grey lines), $1.5B$ (red lines), $2B$ (purple lines), and infinite (black lines). Enforcing periodicity on the covariance matrix is equivalent to restricting its Fourier modes to be harmonics of its wrap-around period. As a result, the covariance matrix is only able to effectively filter these harmonics. For example, when we set periodicity to 10 MHz, our filter only effectively removes the $1/(10\,{\rm MHz}) = 100$ ns tone (dashed black line). When the periodicity is set to 20 MHz, we can remove the 50, 100, and 150 ns tones. When we use 100-MHz bandwidth, tones are spaced by 10 ns. When we set the periodicity to 200 MHz, the spacing between tones drops to ≈5 ns but all tones within the attenuation region are effectively removed due to the finite width of suppression about each tone. The fact that the DFT diagnalized filtering matrix approximately converges to DAYENU at $\gtrsim 2B$ wrap-around indicates that $\sim 4\tau_w B$ modes must be fit in order to achieve similar performance. This can be understood as an approximate manifestation of Nyquist's theorem since we are attempting to describe frequency-limited foregrounds with infinite but highly concentrated support in delay space. Representing such a signal requires at least $\gtrsim 1/2B$ sampling.

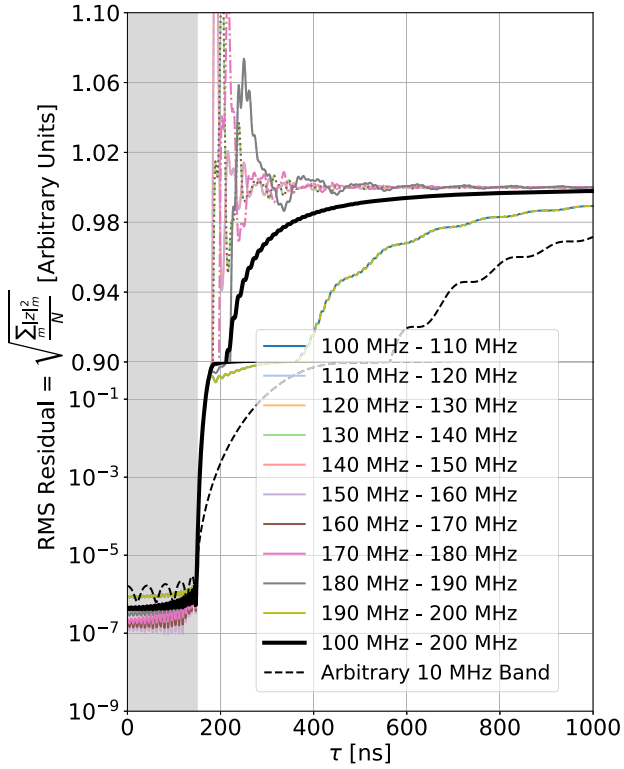

Figure 9. The RMS residual of truncated 10 MHz sub-bands of tones after $\mathbf{R}^{\urcorner}$ is applied to the full 100-MHz band. The degree of signal attenuation is significantly improved over the case where the filter is applied directly to each 10 MHz sub-band after truncation (black dashed line). With the exception of the two outer sub-bands, signal attenuation is below 1 per cent by $\gtrsim$200 ns beyond the filter edge. The edge bands have $\lesssim$10 per cent signal attenuation within 250 ns of the filter edge. Bringing attenuation below $\lesssim$1 per cent brings it within the expected sample variance error bars of spherically average power spectra. Bringing this attenuation below 10 per cent brings it below the expected sample variance of per-baseline power spectra (Lanman & Pober 2019).

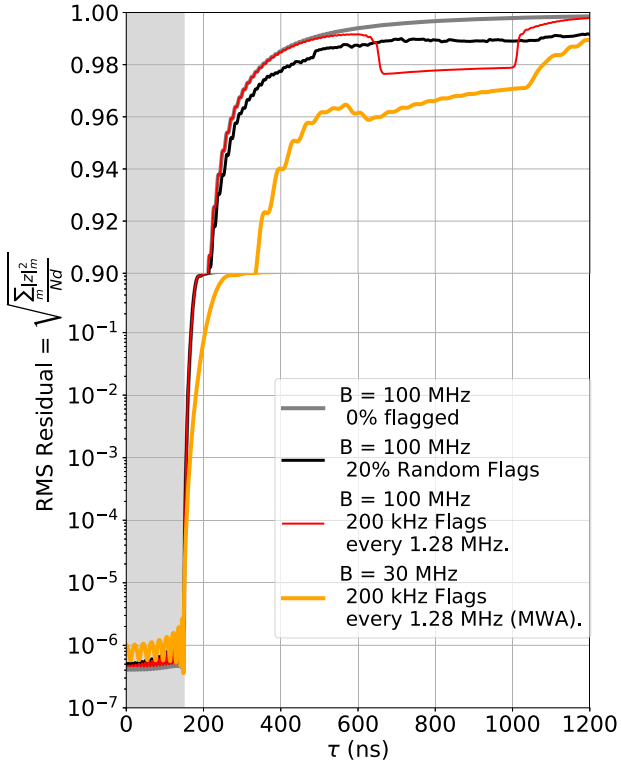

Figure 10. RMS residuals of $\mathbf{R}^{\urcorner}$ for tones filtered with various flagging patterns in data sampled every 100 kHz. We compare no flags over 100 MHz (black line) 200 randomly flagged channels (grey line) and 200 kHz of flagging every 1.28 MHz (red line) – similar to what is typically performed on the MWA. Since the MWA only observes 30 MHz simultaneously, we also show 200 kHz flags every 1.28 MHz (gold line). Random flagging increases attenuation by a percent or so. MWA-like flagging results in ≈2 per cent attenuation over most delays.

beyond the filter edge. In the outer 10 MHz bands, 10 per cent loss is still achieved by 150–200 ns off the filter edge. In light of these results, we recommend sub-band power spectrum estimates be obtained from data on which DAYENU is applied over as wide a band as possible and then truncated.

3.8 Flagged channels

In real life, some fraction of interferometric channels are contaminated by RFI and must be discarded. Thus, it is necessary for DAYENU to work robustly on data that is not evenly sampled. We investigate the impact of RFI flagging by inspecting RMS residuals from applying the psuedo-inverse of $\mathbf{C}^{\urcorner}$ where rows and columns corresponding to flagged channels are set to zero. We explore two different scenarios over 100 MHz of bandwidth. One in which 20 per cent of channels are flagged randomly and one in which 200 kHz flags are applied every 1.28 MHz; similar to what must be performed on the MWA (Dillon et al. 2015; Beardsley et al. 2016; Ewall-Wice et al. 2016b; Barry et al. 2019) (Fig. 10). Since the MWA records ≈30 MHz simultaneously, we also show the RMS residual of $\mathbf{R}^{\urcorner}$ with 200 kHz flags every 1.28 over 30 MHz.

With 200 100 kHz channels flagged randomly over 100 MHz, we find that attenuation beyond the filter width increases by approximately 1 per cent out to large delays. The presence of periodic flags results in the flagging attenuation being concentrated in a concentrated region centred ≈781 ns, the delay of the 1.28 MHz flag periodicity. Outside of this region, the attenuation is negligible but within this region it exceeds 2 per cent, in excess of the average 1 per cent induced by randomized flagging.

3.9 DAYENUREST

By subtracting foregrounds with a matrix multiplication, DAYENU accomplishes one of the primary objectives of the iterative CLEAN filter (Parsons et al. 2012b). $\mathbf{z} = \mathbf{R}^{\urcorner}\mathbf{x}$ is equivalent to the residual after CLEAN is applied. The second goal of CLEAN is to smoothly interpolate (restore) the subtracted foregrounds by adding back their CLEAN components; interpolating the foregrounds over flagged channel gaps with DFT modes. We can isolate the foregrounds subtracted by $\mathbf{R}^{\urcorner}$ with the matrix operation $(\mathbf{I} - \mathbf{R}^{\urcorner})$ and fit them to $N_{\rm DPSS}$ DPSS modes. DPSS vectors are eigenvectors of the foreground component of $\mathbf{C}^{\urcorner}$ so we can approximate our foregrounds with the DPSS vectors with eigenvalues above some small number relative to the largest eigenvalues. We choose a cut-off of 10^{-12} the largest eigenvalue that ensures that foreground modes are subtracted to a level of $\lesssim 10^{-6}$.

Fitting and interpolating with our $N_{\rm DPSS}$ modes can be achieved applying the linear least-squares solution matrix to $(\mathbf{I} - \mathbf{R}^{\urcorner})$

$$\mathcal{A} = \mathbf{A}[\mathbf{A}^T\mathbf{w}\mathbf{A}]^{-1}\mathbf{A}^T\mathbf{w}, \quad (32)$$

where $\mathbf{A}$ is an $N_{\rm d} \times N_{\rm DPSS}$ matrix

$$A_{m\alpha} = u_m^{(\alpha)}(N_{\rm d}, \tau_w), \tag{33}$$

where $u_m^{(\alpha)}(N_{\rm d}, \tau_w)$ is the mth element of the αth DPSS vector of length $N_{\rm d}$ that diagnalizes the $N_{\rm d} \times N_{\rm d}$ matrix $S_{mn}(N_{\rm d}, \tau_w) = (2\tau_w \Delta\nu){\rm Sinc}[2\pi\tau_w(\nu_m - \nu_n)]$ and $\mathbf{w}$ is a diagonal matrix set to unity at unflagged channels and zero at flagged channels. Applying $\mathcal{A}$ to $(\mathbf{I} - \mathbf{R}^{\urcorner})$ provides us with DPSS interpolated CLEAN components. Adding these CLEAN components to the residual gives us a linear REST (restoration) matrix which both filters the data and interpolates the subtracted foregrounds.

$$\mathbf{R}^{\rm REST} = \mathbf{A}[\mathbf{A}^T\mathbf{w}\mathbf{A}]^{-1}\mathbf{A}^T\mathbf{w}(\mathbf{I} - \mathbf{R}^{\urcorner}) + \mathbf{R}^{\urcorner}. \tag{34}$$

We can understand the first term of equation (3.9) as follows. First $(\mathbf{I} - \mathbf{R}^{\urcorner})$ is applied which effectively filters out all small-scale structure dominated by the 21-cm signal and contains RFI flagging gaps. Next, $\mathbf{A}^T\mathbf{w}$ transforms the flagged data into the DPSS basis. Mode-mixing between the DPSS coefficients, due to flagged channels, is undone by applying $[\mathbf{A}^T\mathbf{w}\mathbf{A}]^{-1}$ and a final application of $\mathbf{A}$ transforms back into frequency space. Thus, the total action of the first term is the interpolation over flagged channels with fitted smooth DPSS modes. The second term of equation (3.9) isolates the fine-frequency components of the signal including noise and the 21-cm signal itself.

In Section 4, we will demonstrate the performance of DAYENUREST on realistic foreground and signal simulations.

4 VALIDATION WITH REALISTIC SIMULATIONS

In the last section, we tried to understand how demixing and filtering were limited by non-idealities of the signal covariance matrix. To this end, we simulated Gaussian realizations of a simplified foreground model with no consideration of antenna chromaticity or reference to an actual sky with spectral slope. In addition, the dynamic range that we assumed between foregrounds and 21 cm (eight orders of magnitude in the power spectrum), was somewhat less than what is expected for many models. In this section, we validate DAYENU by applying it to more realistic simulated visibilities.

4.1 Simulation description

In this section, we use simulated HERA visibilities (Appendix A, Kern et al. 2019) to validate filtering with $\mathbf{R}^{\urcorner}$ along with the overall impact of this filtering on power-spectrum statistics. We construct our simulations using the HERALVIS software (Lanman & Kern 2019), which integrates the visibility equation using a HEALPIX representation of the sky (Górski et al. 2005). The simulations use the Global Sky Model (GSM; de Oliveira-Costa et al. 2008) for the foreground model, and a flat-spectrum, uncorrelated random Gaussian field as the EoR model with a variance of 25 mK^2.

They also use a simplified model of the HERA primary beam in instrumental XX and YY polarization, assuming minimal frequency structure in the sidelobes of the beam. Specifically, the beam is low-pass filtered across frequency at every HEALPIX pixel to reject structures for $|\tau| > 250$ ns. For this work this is likely an inconsequential feature of the simulations, as it sets at which delay the foreground power dips below the EoR signal, which is not something that our analysis is sensitive to (Fagnoni et al. 2019). The simulations span 8 h of local sidereal time (LST) and have a frequency coverage from 120 to 180 MHz in 256 channels leading to a 235 kHz channelization. We refer the reader to Lanman et al. (2019) for more details on the HERALVIS package and (Kern et al. 2019) for further information on the simulated data products. RDI plays a major role in setting the efficacy of these techniques. In this section, we use flagging masks representative of the RFI environment for HERA's first observing season (Kerrigan et al. 2019; Kern et al. 2020).

4.2 Validating DAYENU and DAYENUREST as visibility filters

Aside from being used as a filtering matrix in the final calculation of $\hat{p}_\alpha$, DAYENU can readily be employed in sandbox-type data analyses assessing the level of spectral structures in individual visibilities, data-cubes, and other products. In this section, we compare its efficacy to CLEAN filtering which is often used to a similar end. To do so, we inspect the performance of the direct application of DAYENU and DAYENUREST to our simulated visibilities, and compare our results to CLEAN. In the literature (e.g. (Kern et al. 2019)), CLEANing is performed on the visibility after zero-padding by $N_{\rm d}$ channels on either side (For these simulations $N_{\rm d} = 256$) and taper-filtering with a Tukey window with $\alpha = 0.15$. Zero-padding is performed to give CLEAN a larger number of Fourier modes to work with; allowing it to fit the same aharmonic delays that are absent from an $N_{\rm d}$ DFT. We perform CLEANing over ± 150 ns in delay-space. Each iteration of CLEAN finds the peak power of the data in delay-space and subtracts the peak power times 0.1 (gain) times a flagging kernel centred at the peak delay until the RMS residual changes with each iteration by less than some fraction of the RMS of the original visibilities. The tolerance parameter can be set as low as we want to obtain some arbitrary degree of foreground subtraction. In practice, the choice of tolerance depends on the constraints of computational resources. We adopt 10^{-9} that is currently being used in the HERA analysis pipeline. In addition, for $N_{\rm d} = 256$, CLEANing a single baseline on a single time to 10^{-9} tolerance has a similar runtime (within an order of magnitude) of computing the psuedo-inverse of $\mathbf{C}^{\urcorner}$ to obtain $\mathbf{R}^{\urcorner}$.

For DAYENUREST, we limit the set of DPSS vectors to those with eigenvalues of $\mathbf{L}$ greater then 10^{-12}. As we stated in Section 3.3, the maximum eigenvalue of $\mathbf{L}$ is close to unity. We compare the sum of clean residuals and clean components, which interpolate over flagged channel gaps (center, Fig. 11), to DAYENURESTd simulations (right, Fig. 11). At large scales, our linear cleaning and interpolation technique performs just as well as CLEAN in reproducing macroscopic foreground features. In order to understand the low-level disagreements between the two, we inspect their residuals.

We compare the residuals from CLEAN and DAYENUREST (Fig. 12). For CLEAN, we refer to residuals as what is left in the data after iteratively subtracting all CLEAN-components and for DAYENU and DAYENUREST, as in the previous sections, residuals refer to the data after applying $\mathbf{R}^{\urcorner}$. Note that the residuals for DAYENU and DAYENUREST are identical by the definition of DAYENUREST (equation 3.9). In Fig. 12, DAYENU and DAYENUREST subtract the foregrounds to below the 21-cm level (right-hand panel) while CLEAN leaves significant residuals (center right panel). To understand the impact of flagging, we also inspect the residuals of CLEAN with no flagging (center left panel). The CLEAN residuals are nearly identical whether or not flagging is present. It follows that flagging alone does not impact the absolute level of residuals left after CLEANing. If these residuals intrinsically stay within the wedge, they will not have an impact on our ability to detect 21 cm outside of the wedge. However, the presence of flagged channels will cause the residuals to enter the EoR window at a level that depends on the flagging.

In Fig. 13, we compare the Blackman–Harris taper-filtered delay-transform of DAYENUREST and CLEAN filtered data with and without flagging across three different bands. For DAYENUREST filtered data

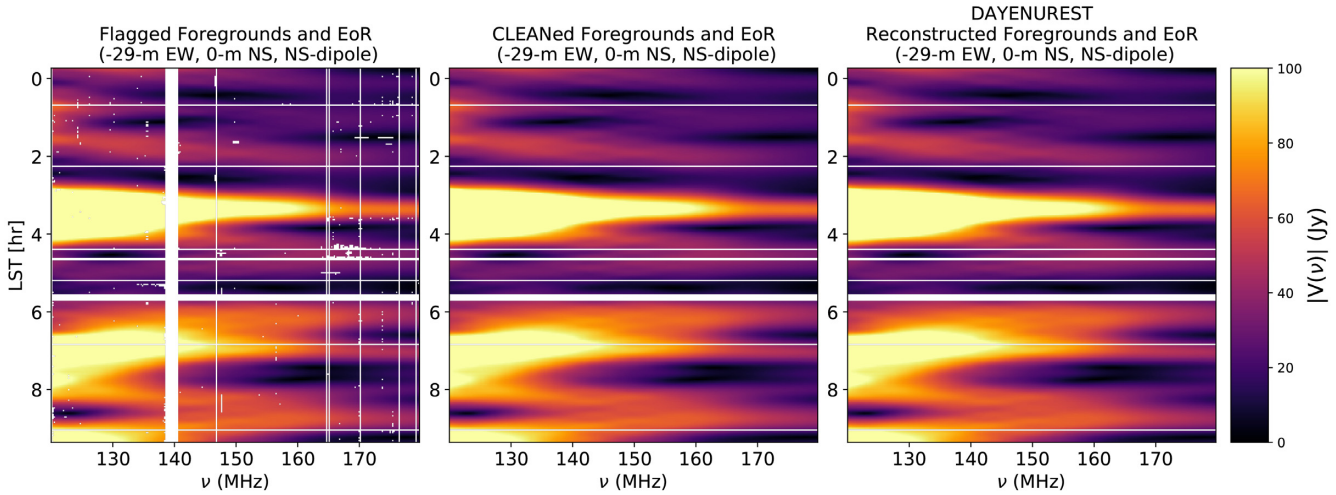

Figure 11. Left: A simulated visibility including modelled foregrounds and 21-cm fluctuations with gaps at the locations of frequency dependent RFI flags. Center: Simulated foregrounds and EoR after low-delay frequency interpolation with the CLEAN algorithm. Right: Simulated foregrounds and EoR after low-delay frequency interpolation with DAYENUREST. At the macro-scale, linear in-painting delivers qualitatively similar results to iterative CLEANING. The low-level inconsistences between foreground interpolation by CLEAN and DAYENUREST are best understood by inspecting the residuals left over after subtracting these foreground models (Figs 12 and 13).

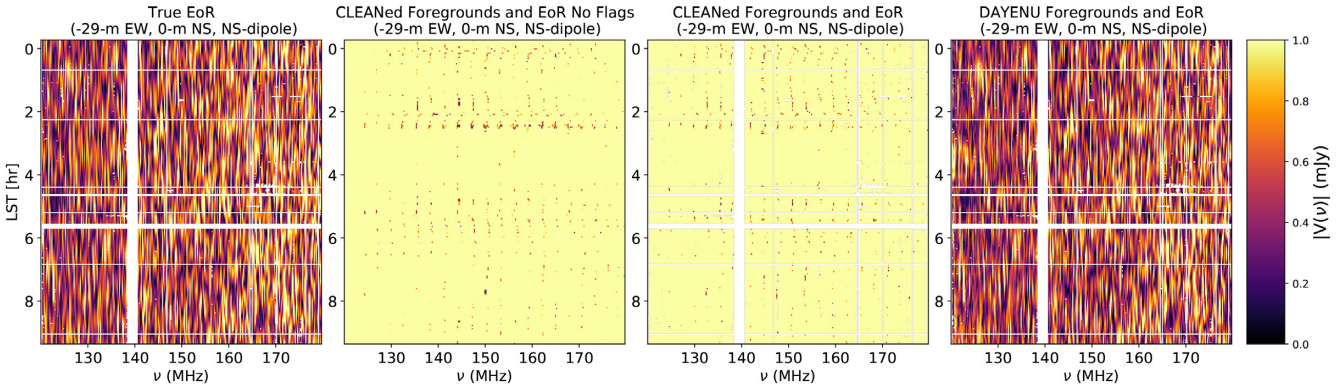

Figure 12. Left: an injected mock EoR signal. Center left: Residuals after filtering using the CLEAN algorithm with no flagging. Center right: Residuals after foreground filtering using the CLEAN algorithm with flagging. The level of real-space CLEAN residuals is roughly independent of flagging. Although the CLEAN residuals exceed the 21-cm signal, as long as these residuals are spectrally smooth, they are not an obstacle to detecting 21 cm in Fourier space. The presence of flagging and residuals presents complications (as we see below in Fig. 13). Right: Residuals after foreground filtering using our linear filter. EoR fluctuations remain primarily intact while foregrounds have been completely eliminated.

refer to the data after the application of $\mathbf{R}^{\mathrm{REST}}$. For CLEAN filtered refers to CLEAN residuals plus the interpolating CLEAN components. Our three bands are as follows. First, the entire 120–180 MHz band. Secondly, a 120–138 MHz band below ORBCOMM which is heavily flagged, and thirdly 141–180 MHz above ORBCOMM with roughly twice the bandwidth as below. With no RFI flagging, CLEAN and DAYENUREST perform similarly well as can be seen by comparing the red-solid and grey-solid lines in Fig. 13. Unfortunately, the presence of RFI flags causes significant bleed of the CLEAN filtered data outside of the wedge and is especially bad when the DFT band includes ORBCOMM at 137 MHz. We also plot the residuals of CLEAN and DAYENUREST as dashed lines. The maximum low-delay level of CLEAN residuals is practically the same with and without flags. The presence of flags causes these residuals to bleed to high delays at levels much larger then 21 cm. Since the level of these bleeding residuals agrees with the level of the total filtered data, we conclude that the structures in CLEAN residuals introduced by flagging are to blame for high-delay contamination in the CLEAN filtered visibilities. Even without ORBCOMM, leakage of CLEAN residuals exceeds our injected 21-cm signal by a factor of a few. DAYENUREST (red-solid line) successfully removes foregrounds below the level of the 21-cm signal (black dotted line) in all cases. The relatively narrow bandwidth below ORBCOMM, presents a potential challenge since the central foregound lobe extends to $k_{\parallel} \approx 0.2\,h\,\mathrm{Mpc}^{-1}$. Losing $k_{\parallel} \lesssim 0.2\,h\,\mathrm{Mpc}^{-1}$ to foregrounds has a significant impact on science returns (Pober et al. 2014; Ewall-Wice et al. 2016a, c). In Section 4.3, we investigate whether the central foreground lobe is actually a fundamental limitation.

Over 256 channels, CLEAN's runtime per integration is also significantly larger than DAYENUREST's. With our adopted parameters, on a laptop with a 2.4 GHz i5 processor, computing $\mathbf{R}^{\neg}$ for each unique flagging pattern and set of filter-widths, centres, and suppression factors takes roughly 0.24 s while filtering a baseline at a single time with a cached filter matrix takes approximately 0.003 s. In comparison, the time for CLEAN to run on each baseline-time is 0.8 s and there is no possibility of speeding things up through caching.

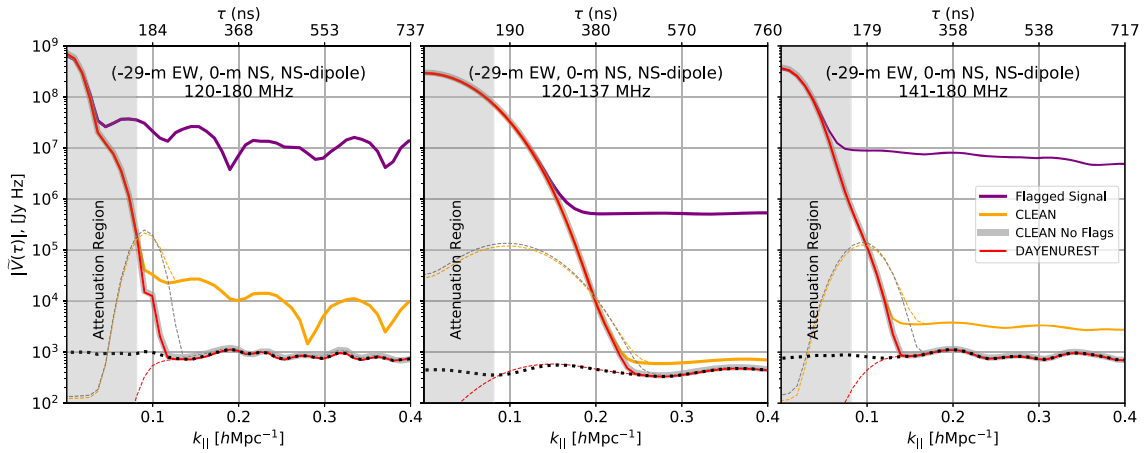

Figure 13. Time averages over 8 h of LST of the absolute value of delay-transformed visibilities in Fig. 11, tapered by a 7-term Blackman–Harris window. Left: 120–180 MHz (all 256 channels), centre: 120–137 MHz (below ORBCOMM), and right: 141–179 MHz (above ORBCOMM). Solid lines represent filtered and restored foregrounds and thin dashed lines show residuals. We show the attenuation of our CLEAN andDAYENU filters as a grey-shaded region. Over all bands, ringing from RFI flags causes the unfiltered foregrounds (purple lines) to completely mask the 21-cm signal (black-dotted lines). Pealing and in-painting foregrounds using the CLEAN algorithm with a tolerance of 10^{-9} leaves significant residuals that exceed the 21-cm signal in all studied bands and are especially problematic when the FT window includes the heavily flagged ORBCOMM frequencies (≈137 MHz). DAYENUREST (dashed line) subtracts foregrounds far below the 21-cm level, allowing for an unbiased estimate of 21-cm emission outside of the central foreground lobe.

Before we move on to power spectra, it is worth noting that although we have focused filtering visibilities, $\mathbf{R}^{\urcorner}$ can just as easily be used to foreground-filter gridded visibilities by applying $\mathbf{R}^{\urcorner}$ along the frequency axis of each *uv* cell. In this situation, one would set τ_w to include not only the intrinsic chromaticity of the antenna and the wedge in the *uv* cell but also to include any additional spectral structure that might be introduced by gridding. We leave the question of how much one would need to increase τ_w for different gridding strategies to future work.

4.3 Power spectra

We now explore the impact that various choices of $\mathbf{R}$ have on the final power spectrum when we use identity normalization $\mathbf{M} \propto \mathbf{M}_{\rm ID}$. We calculate a normalized $\hat{\mathbf{p}}$ from 42 channels between 145 and 155 MHz; corresponding to a redshift interval of $\Delta z \approx 0.5$ for the following choices of $\mathbf{R}$.

(i) **Blackman–Harris:** We use an apodization filter with the diagonal set equal to a 7-term Blackman–Harris taper function $\mathbf{R} = \mathbf{R}^{\rm BH}$. To obtain a noise-equivalent bandwidth of 10 MHz, we extend the spectral window to 96 channels (22.5 MHz).

(ii) **No flags:** A scenario for reference. The same as simple delay-spectrum but with no RFI flagging. In this scenario, we also have $\mathbf{R} = \mathbf{R}^{\mathbf{BH}}$.

(iii) DAYENU **Narrowband:** Apply $\mathbf{R}^{\urcorner}$ with $\epsilon = 10^{-9}$ and $\tau_w = 150$ ns across the same bandwidth as the Fourier transform (42 channels – 10 MHz; $\mathbf{R}^{\urcorner}$). We do not use a taper in the Fourier transform. Thus $\mathbf{R} = \mathbf{R}^{\urcorner}$.

DAYENU **Restored:** Perform linear inpainting of foregrounds using DAYENUREST with a 150 ns attenuation region and in-painting modes spaced by 44.44 ns ($\mathbf{R}^{\rm REST}$). An identical Blackman–Harris tapered Fourier transform as our Blackman–Harris scenario is used to estimate bandpowers from the filtered data. Thus $\mathbf{R} = \mathbf{R}^{\mathbf{BH}}\mathbf{R}^{\mathbf{REST}}$.

(iv) DAYENU **Extended filter:** We perform filtering across the entire 60-MHz band with $\mathbf{R}^{\urcorner}$ before truncating and performing a DFT across the central 10 MHz, $\mathbf{R} = \mathbf{R}^{\urcorner}$.

In all cases, we use $\mathbf{Q}_\alpha = \mathbf{Q}_\alpha{}^{\mathbf{DFT}}$. In order to convert our power spectra from visibility to cosmological units, we multiply $\mathbf{M}_{\rm ID}$ by a constant

$$\mathbf{M} = S \times \mathbf{M}_{\rm ID}, \tag{35}$$

where

$$S = \left(\frac{\lambda^2}{2k_B}\right)^2 \frac{X^2 Y}{N_{\rm d}^2 \Omega_{pp} B}, \tag{36}$$

Ω_{pp} is the solid angle integral of the primary beam squared and averaged over our band of interest, $Y = dr_{||}/d\nu$, $X = dr_{\perp}/d\theta$, λ is the average observation wavelength, and k_B is the Boltzmann constant. We refer the reader to Morales & Hewitt (2004), Parsons et al. (2012a, 2014) for more the full expressions of these constants and their derivations. We estimate power spectra from 8 h of LST by computing an independent $\hat{\boldsymbol{p}}$ every 30.6 s and incoherently averaging. Our bandpower estimates appear in Fig. 14 along estimates of vertical and horizontal 68 per cent confidence errorbars. We derive these confidence intervals from estimates of the bandpower covariances $\hat{\mathbf{\Sigma}}$ and window-functions $\hat{\mathbf{W}}$. Before we discuss the results in this plot we first describe our calculations $\hat{\mathbf{\Sigma}}$ (Section 4.3.1) and $\hat{\mathbf{W}}$ (Section 4.3.2).

4.3.1 Error bars

To calculate $\hat{\sigma}^{\hat{p}}_\alpha$, the standard deviation of our αth bandpower after incoherent averaging, we first calculate $\hat{\sigma}^0_\alpha \equiv \sqrt{\hat{\Sigma}_{\alpha\alpha}}$ by empirically computing the covariance of $\hat{p}$ across all LSTs. We show our estimates of $\hat{\Sigma}$ in Fig. 15. To account for the reduction in errors that occurs from incoherently averaging over the independent realizations of foregrounds and 21-cm fluctuations in the sky, we use the equation

$$\hat{\sigma}^{\hat{p}}_\alpha = \hat{\sigma}^0_\alpha \sqrt{\frac{\mathrm{FWHM}^\alpha_c}{T}}, \tag{37}$$

where FWHM^α_c is the full-width half-maximum in time of the correlation between the $\alpha^{\rm th}$ bandpower and itself $\hat{\Sigma}_{\alpha\alpha}(\Delta t)$ and T is the total amount of time over which LSTs are averaged (8.5 h). We

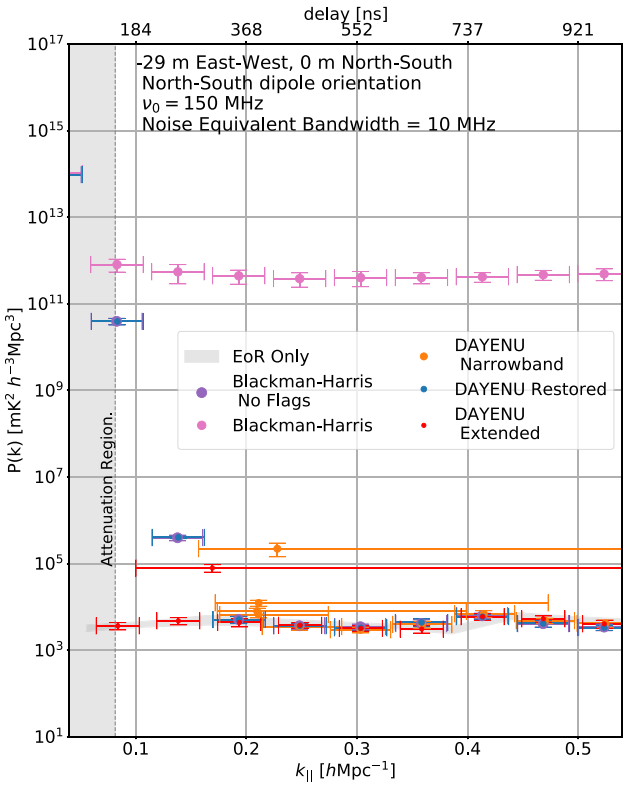

Figure 14. Power spectra estimated from a −29 m east–west oriented baseline over 10 MHz noise equivalent bandwidth centred at 150 MHz and 8 h of LST. Vertical error bars are 68 per cent confidence regions computed from the diagonal of $\hat{\mathbf{\Sigma}}$ and arise from the sample-variance in 8 h of sky observations (Section 4.3.1). Horizontal errorbars are the 68 per cent confidence intervals derived from estimates of the window-function matrix $\hat{\mathbf{W}}$ (Section 4.3.2) and points are plotted at 50 per cent point of each $\hat{\mathbf{W}}$ row. With only a Blackman–Harris apodization filter applied, power spectrum estimates are heavily contaminated by flagging sidelobes of the foregrounds (pink points). Filtering with DAYENUREST and a Blackman–Harris both interpolates the flagged channels and removes power associated with the sharp edges of our finite sample bandwidth (blue points), resulting in a measurement that is in general agreement with an unflagged Blackman–Harris tapered DFT (purple points). Tapered DFT methods that leave the foregrounds in must contend with those foreground's sidelobes. Over 10 MHz NEB, these sidelobes extend to $\sim 0.2\,h\,\mathrm{Mpc}^{-1}$, rendering measurements of larger scale modes highly contaminated by foreground bias. DAYENU is a filter that targets and removes foregrounds. But unintentional attenuation of the signal also occurs beyond the edge of the attenuation region (vertical grey filled region) specified by τ_w. If we apply DAYENU over 10 MHz then this attenuation is significant in our single baseline power spectrum out to $0.2\,h\,\mathrm{Mpc}^{-1}$ (orange points). ApplyingDAYENU across 60 MHz before estimating our bandpowers from the central 10 MHz sub-band allows us to measure bandpowers down to $\sim 0.1\,h\,\mathrm{Mpc}^{-1}$ with relatively small bias which can be further mitigated using more sophisticated normalization.

compute bandpower time-correlations using

$$\hat{\Sigma}_{\alpha\alpha}(\Delta t) = \frac{1}{N_t}\sum_t \hat{p}_\alpha(t+\Delta t)\hat{p}^*_\alpha(t), \tag{38}$$

where N_t is the number of times and $\hat{p}_\alpha(t)$ is the bandpower estimate at each time-step. In our case, $N_t = 1000$. We find the full-width half-maximum of $\hat{\Sigma}_{\alpha\alpha}(\Delta t)$ using the method `scipy.signal.find_peaks`. In Fig. 14, we show the averaged bandpowers and 2σ error bars. Since our simulation does not include noise, the errors are purely sourced by sample variance in the foregrounds and signal.

4.3.2 *Window matrices*

We estimate window matrices using the equation

$$\hat{\mathbf{W}} = \mathbf{M}\hat{\mathbf{H}}, \tag{39}$$

where

$$\hat{H}_{\alpha\beta} = \frac{1}{2}\mathrm{tr}(\mathbf{R}^\dagger \mathbf{Q}_\alpha \mathbf{R}\widehat{\mathbf{C}}_{,\beta}). \tag{40}$$

In practice we do not necessarily have $\hat{\mathbf{H}} = \mathbf{H}$ since we do not know the a priori actual bandpowers of the signal in question and are instead forced to guess some $\widehat{\mathbf{C}}_{,\beta}$. While we technically do potentially have the ability to calculate true bandpowers for our simulated visibilities, we defer an exploration of the consequences of not using true bandpowers to compute $\mathbf{H}$ for paper II. In this paper, we adopt the standard DFT bandpower assumption so that $\widehat{\mathbf{C}}_{,\beta} = \widehat{\mathbf{C}}_{,\beta}^{\mathrm{DFT}}$.

We show $\hat{\mathbf{W}}$ for our various $\mathbf{R}$ choices, averaged over all time-samples, in Fig. 16. Our window functions for the Delay Spectrum and DAYENU Restored are very close to each-other outside of the filtering region where they are narrowly peaked but level off at ≈ -35 dB. We also plot every fourth row of $\hat{\mathbf{W}}$ for an estimator with no flagging and a Blackman–Harris apodization filter in Fig. 16. Since these window functions continue to descend below −35 dB, we conclude that the −35 dB floor in most $\hat{\mathbf{W}}$ rows is a consequence of flags. In our Blackman–Harris estimator, these −35 dB sidelobes extend from bandpower estimates inside of the attenuation region just as much as bandpower estimates outside of the attenuation region. If no foregrounds are subtracted, bandpower estimates inside of the attenuation region are heavily contaminated by foregrounds, causing the significant contamination across all bandpowers that we observe in the Blackman–Harris model (pink points) in Fig 14. Since the vast majority of power within the filtering region is sourced by interpolated and effectively unflagged DPSS modes, the DAYENU Restored filter removes the components of sidelobes of bandpowers centred outside of the attenuation region that overlap with the attenuation region. This effectively breaks the coupling of modes outside the attenuation region with the foregrounds. The DAYENU Narrowband filter suppresses the coupling of all bandpower estimates with delays inside of the attenuation region and as a consequence, many of the rows of $\hat{\mathbf{W}}$ that would typically be centred inside of the attenuation region are now centred at its edge at $k_\parallel \approx 0.2\,h$ Mpc and preventing us from effectively measuring cosmological modes below this value. By extending the filtering bandwidth from 10 to 60 MHz our DAYENU Extended filter reduces the width of the attenuation region to $\approx 0.1\,h\,\mathrm{Mpc}^{-1}$ and allowing for significant improvements in our ability to detect and interpret 21-cm fluctuations.

4.3.3 *Power-spectrum results*

Having explained the source of our vertical and horizontal 68 per cent confidence regions, we discuss the results of Fig. 14. The presence of RFI gaps introduces window-function sidelobes at the −35 dB level (Fig. 16). Thus, if our $\mathbf{R}$ filter does not attenuate foregrounds before applying $\mathbf{Q}_\alpha^{\mathrm{DFT}}$, all bandpowers will be heavily contaminated by foregrounds. This is indeed the case for our Blackman–Harris model (pink points). If no flags are present, these flagging sidelobes do not exist and our estimator eventually recovers 21 cm. However, the

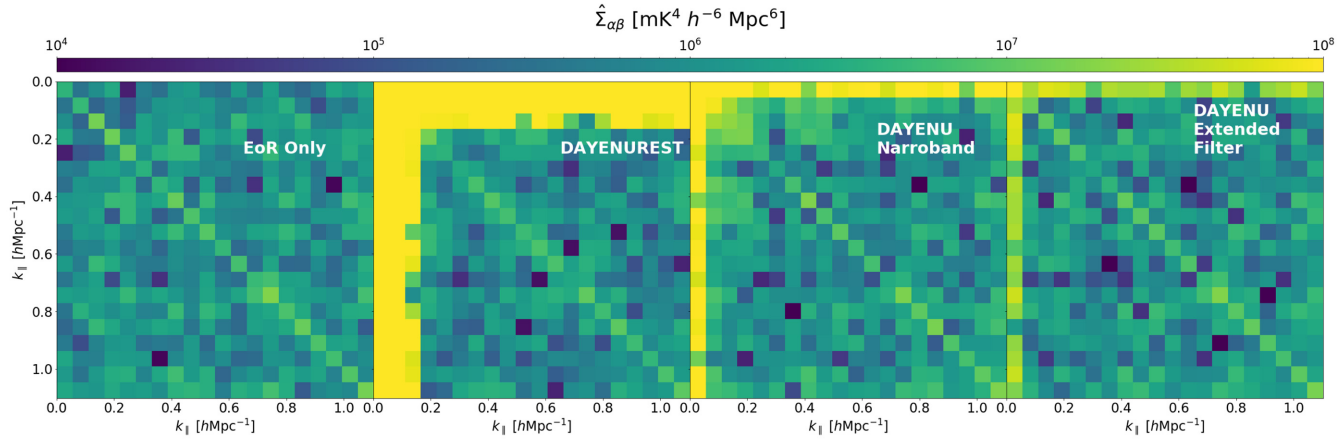

Figure 15. The covariance matrices of $\hat{\boldsymbol{p}}$ $\hat{\boldsymbol{\Sigma}}$ from which the errorbars in Fig. 14 are derived. Left: As a result of flagging, and not attempt to decorrelate, power-spectrum errors for the DFT of EoR simulated data are highly correlated. Center left: Errors from DAYENUREST which restores foregrounds using linear interpolation with DPSSs and as a result, requires a taper-filtered FT over a larger bandwidth. Error bars are very large below $k_\parallel \lesssim 0.2\,h\,\mathrm{Mpc}^{-1}$ but outside of the foreground region, they are somewhat less correlated then the EoR only panel. This is in part because of the larger DFT and lower sidelobes from a Blackman–Harris. Centre right: $\hat{\boldsymbol{\Sigma}}$ for DAYENU applied over the same 10-MHz bandwidth of the DFT. Large foreground errors are now contained within the DC bin but significant error correlations exist below $k_\parallel \lesssim 0.2\,h\,\mathrm{Mpc}^{-1}$. **Right:** $\hat{\boldsymbol{\Sigma}}$ for our DAYENU Extended filtering estimator. Correlations between large $k_\parallel$ modes are similar to the EoR-only and *DAYENU* panels. However, the strong correlations at $k_\parallel \lesssim 0.2\,h\,\mathrm{Mpc}^{-1}$ that exist when DAYENU is applied over a smaller bandwidth have been greatly reduced, as have the foreground errors in the $k_\parallel = 0\,h\,\mathrm{Mpc}^{-1}$ bin.

smallest $k_\parallel$ that we can access is limited by the Blackman–Harris sidelobes of foregrounds which extend to $k_\parallel \sim 0.2\,h\,\mathrm{Mpc}^{-1}$. The same is true for the DAYENU Restored scenario (blue points). The primary accomplishment of foreground interpolation is to remove the bleed from flagging gaps but we must still contend with the Blackman–Harris sidelobes. DAYENU Narrowband (orange points) eliminates foregrounds but also severely attenuates signal out to $\approx 0.2\,h\,\mathrm{Mpc}^{-1}$. Thus, we are still restricted to $k_\parallel \gtrsim 0.2\,h\,\mathrm{Mpc}^{-1}$ and samples that would otherwise be foreground contaminated at smaller $k_\parallel$ are instead primarily contributed to by power just outside the attenuation region, leading to the handful of points with very large horizontal error bars piled up at $k_\parallel \approx 0.2\,h\,\mathrm{Mpc}^{-1}$. By using a larger bandwidth in the filtering step, DAYENU Extended reduces the region of excessive attenuation down to $\lesssim 0.1\,h\,\mathrm{Mpc}^{-1}$ (red points). Hence, by filtering foreground selectively, we can access significantly larger co-moving scales then if we only use apodization tapers. From Fig. 4, we know that our bandpowers are biased low at the 1–10 per cent level – something that is technically not significantly detected in our single-baseline analysis due to sample variance errors. However, this bias can have implications for more sensitive spherically binned power spectra.

5 CONCLUSIONS

In this paper, we introduced a new method for subtracting foregrounds with a highly approximated inverse covariance filter that we call DAYENU. With no flagging, DAYENU effectively filters foregrounds using DPSSs which are a set of sequences that maximize power concentration within the wedge. Unlike apodization filters, which subtract power equally from foregrounds and signal, DAYENU targets and subtracts low-delay foregrounds with minimal impact on high delay signal and noise. DAYENU avoids the band edge signal attenuation that is a feature of multiplicative taper filters. DAYENU is fast, only requiring that one take the psuedo-inverse of a modestly sized analytical covariance for each baseline length and unique flagging pattern while its linearity allows us to propagate its effect into error estimates and other statistical calculations. We have tested DAYENU on simulated visibilities, but in principal it can also filter foregrounds from gridded *uv* data by applying it to each *uv* cell instead of each baseline provided that τ_w is increased sufficiently to include gridding artefacts. Applying DAYENU to realistic simulations, we have learned the following:

(i) DAYENU is effective at subtracting delay-limited foregrounds at the $\lesssim 10^{-6}$ level, even in the presence of significant flagging (Figs 3 and 12). If applied across an $\approx$100 MHz band, signal attenuation is kept below $\approx$1 per cent beyond 300 ns of the delay-space filter edge. This attenuation can be corrected further in the power-spectrum normalization step. DAYENU's efficacy over filtering with a DFT arises from the fact that, unlike the DFT, it down-weights foreground wedge structures that are not harmonics of B^{-1}.

(ii) A combination of DAYENU and least-squares fitting of DPSSs (DAYENUREST) is a fast, linear alternative to the iterative CLEAN algorithm whose residuals are significantly smaller than CLEAN 's given similar computing times (Figs 11 and 12).

(iii) Applying DAYENU across an $\sim$60–100 MHz band before estimating bandpowers over the $\sim$10 MHz necessary for stationary 21-cm statistics allows us to access LoS scales of $\lesssim .15\,h\,\mathrm{Mpc}^{-1}$ that, even without flagging, are inaccessible to apodized DFTs of Fig. 14 and Fig. 16.

Our takeaway from examining DAYENU is that in the regime where baselines are short so that their information is mutually independent, an inverse covariance filter that is good enough for us is simply one that captures the large dynamic range between foregrounds and signals over the wedge delays and includes information on the frequency structures in the foreground wedge that are not harmonics of B^{-1}. We have shown that a simple covariance like $\mathbf{R}^\neg$ can be many orders of magnitude different from that of the true data covariance but still serve as a highly effective filter. This bodes well for 21 cm and other intensity mapping applications where the precision characterization of our instruments and foregrounds is difficult.

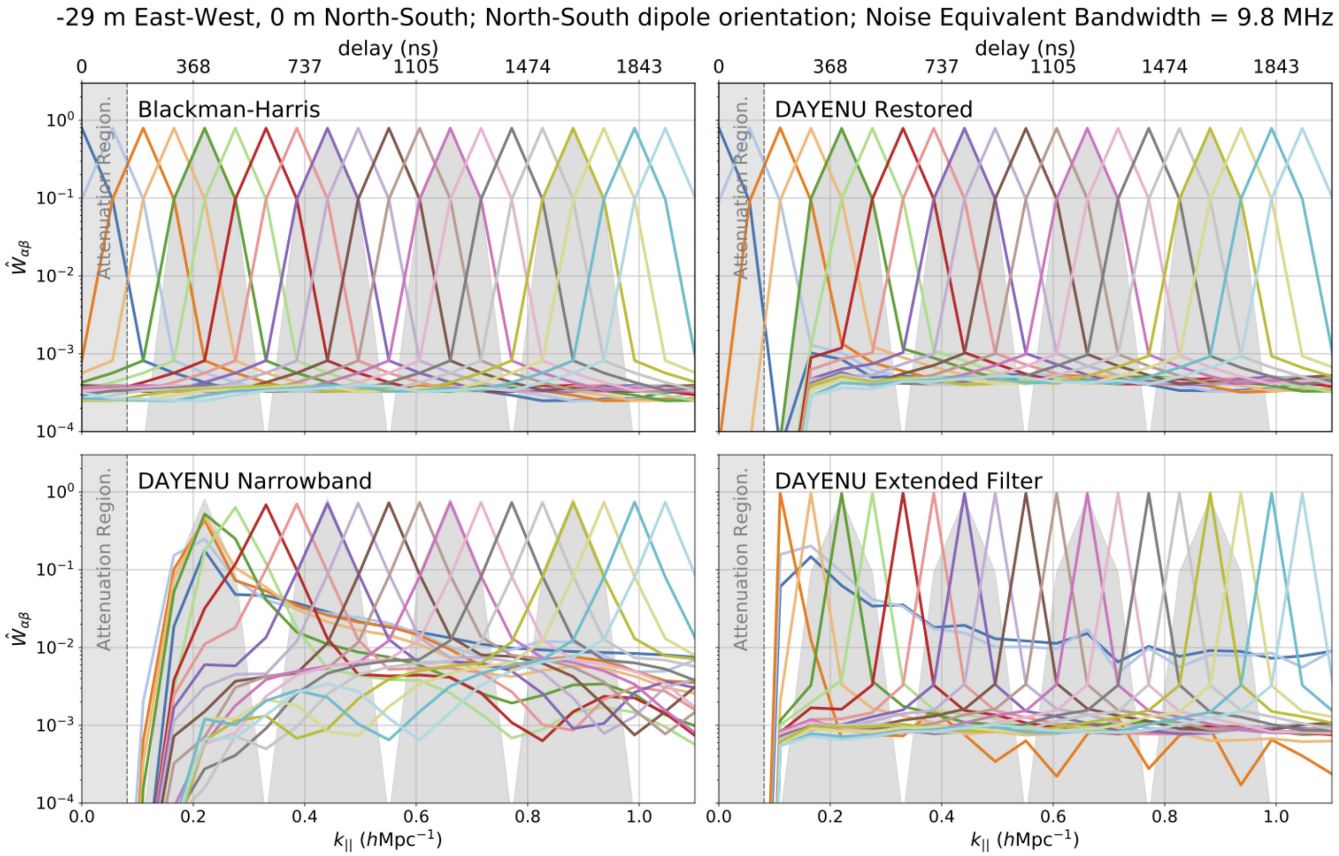

Figure 16. Rows of $\hat{\mathbf{W}}$ for the various choices of $\mathbf{R}$ considered in this paper. Each coloured line is a different row. We also show every fourth row of $\hat{\mathbf{W}}$ for an unflagged Blackman–Harris filtered power spectrum as grey shaded regions. The attenuation set by τ_w in DAYENU is also indicated by a grey shaded region bordered by a dashed line. Top left: Rows of $\hat{\mathbf{W}}$ when only a Blackman–Harris filter is used on the flagged visibilities. Window functions exhibit a floor at ≈ -35 dB arising from the flags. Top right: When we use the DAYENUREST filter, flagging gaps are interpolated over by DPSS vectors that span the attenuation region. This results in the removal of the flagging sidelobes of bandpowers centred within the attenuation region, preventing foreground leakage. Flagging sidelobes remain outside of the attenuation region. Bottom left: Applying DAYENU across a narrow band (10 MHz) removes power within the 150 ns attenuation region along with associated sidelobes, eliminating the problem of foreground-flagging sidelobes contaminating all bandpowers. $\hat{\mathbf{W}}$ rows that would otherwise be centred inside of the attenuation region are now centred outside and have larger sidelobes that extend to larger wavenumbers. This is because these bandpowers had most of their power eliminated by DAYENU but the flagged DFT leaks power back in from high delays. The relatively large amount of unintentional attenuation that accompanies a narrow band filter (see also Fig. 4) prevents us from effectively measuring bandpowers below $k_{||} \lesssim 0.2\ h\,\mathrm{Mpc}^{-1}$. Bottom right: Our DAYENU Extended filter filters over all 60 MHz of before performing a DFT over the same 10 MHz as the DAYENU and DAYENU Restored scenarios. The reduction in unintentional attenuation results in our ability to measure 21 cm fluctuations down to $\sim 0.1\ h\,\mathrm{Mpc}^{-1}$, enhancing our ability to perform sensitive 21-cm measurements.

CODE

An interactive jupyter tutorial on using DAYENU can be found at https://github.com/HERA-Team/uvtools/blob/master/examples/linear_clean_demo.ipynb. DAYENU's source code can be found at https://github.com/HERA-Team/uvtools/blob/master/uvtools/dspec.py

This work made use of the NUMPY (Virtanen et al. 2020), SCIPY (Virtanen et al. 2020), MATPLOTLIB (Hunter 2007), AIPY https://github.com/HERA-Team/aipy, and ASTROPY https://www.astropy.org/ and JUPYTER https://github.com/jupyter/jupyter python libraries along with PYUVDATA (Hazelton et al. 2017) and HEALVIS (Lanman & Kern 2019) PYTHON packages.

ACKNOWLEDGEMENTS

We thank Jacqueline Hewitt, Honggeun Kim, Kevin Bandura, Miguel Morales, Bobby Pascua, Bryna Hazelton, and Ue-Li Pen for helpful discussions. AEW and acknowledges support from the NASA Postdoctoral Program and the Berkeley Center of Cosmological Physics. JSD gratefully acknowledges the support of the NSF AAPF award #1701536. A portion of this work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. AL acknowledges support from the New Frontiers in Research Fund Exploration grant program, a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant and a Discovery Launch Supplement, the Sloan Research Fellowship, as well as the Canadian Institute for Advanced Research (CIFAR) Azrieli Global Scholars program. This material is based upon work supported by the National Science Foundation under grants #1636646 and #1836019 and institutional support from the HERA collaboration partners. This research is funded in part by the Gordon and Betty Moore Foundation. HERA is hosted by the South African Radio Astronomy Observatory, which is a facility of the National Research Foundation, an agency of the Department of Science and Innovation.

REFERENCES

Ali Z. S. et al., 2015, ApJ, 809, 61
Anderson C. J. et al., 2018, MNRAS, 476, 3382
Bandura K., et al., 2014, Proc. SPIE 9145, Ground-based and Airborne Telescopes V, SPIE, Bellingham, WA. p. 914522
Barry N., Beardsley A. P., Byrne R., Hazelton B., Morales M. F., Pober J. C., Sullivan I., 2019, PASA, 36, e026
Beardsley A. P. et al., 2016, ApJ, 833, 102
Blake C., Wall J., 2002, MNRAS, 337, 993
Carroll P. A. et al., 2016, MNRAS, 461, 4151
Chang T.-C., Pen U.-L., Bandura K., Peterson J. B., 2010, Nature, 466, 463
Chapman E. et al., 2012, MNRAS, 423, 2518
Chen X., , 2015, IAU General Assembly. Springer, London, UK, p. 2252187
Cheng C. et al., 2018, ApJ, 868, 26
Datta A., Bowman J. D., Carilli C. L., 2010, ApJ, 724, 526
de Oliveira-Costa A., Tegmark M., Gaensler B. M., Jonas J., Landecker T. L., Reich P., 2008, MNRAS, 388, 247
DeBoer D. R. et al., 2017, PASP, 129, 045001
Di Matteo T., Ciardi B., Miniati F., 2004, MNRAS, 355, 1053
Dillon J. S., Liu A., Tegmark M., 2013, Phys. Rev. D, 87, 043005
Dillon J. S. et al., 2015, Phys. Rev. D, 91, 123011
Eastwood M. W. et al., 2018, AJ, 156, 32
Ellingson S. W., Clarke T. E., Cohen A., Craig J., Kassim N. E., Pihlstrom Y., Rickard L. J., Taylor G. B., 2009, IEEE Proc., 97, 1421
Epstein C., 2007, Introduction to the Mathematics of Medical Imaging, 2nd edn. Society for Industrial and Applied Mathematics, Philadelphia, PA
Ewall-Wice A., Hewitt J., Mesinger A., Dillon J. S., Liu A., Pober J., 2016a, MNRAS, 458, 2710
Ewall-Wice A. et al., 2016b, MNRAS, 460, 4320
Ewall-Wice A. et al., 2016c, ApJ, 831, 196
Fagnoni N. et al., 2020, MNRAS, 500, 1232
Górski K. M., Hivon E., Banday A. J., Wand elt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Hazelton B. J., Jacobs D. C., Pober J. C., Beardsley A. P., 2017, J. Open Source Softw., 2, 140
Högbom J. A., 1974, A&AS, 15, 417
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Jacobs D. C. et al., 2011, ApJ, 734, L34
Jacobs D. C. et al., 2016, ApJ, 825, 114
Jacobs D. C. et al., 2017, PASP, 129, 035002
Kern N. S., Parsons A. R., Dillon J. S., Lanman A. E., Fagnoni N., de Lera Acedo E., 2019, ApJ, 884, 105
Kern N. S. et al., 2020, ApJ, 888, 70
Kerrigan J. et al., 2019, MNRAS, 488, 2605
Kolopanis M. et al., 2019, ApJ, 883, 133
Lanman A. E., Kern N., 2019, , Astrophysics Source Code Library, record ascl:1907.002
Lanman A. E., Pober J. C., 2019, MNRAS, 487, 5840
Lanman A. E., Pober J. C., Kern N. S., de Lera Acedo E., DeBoer D. R., Fagnoni N., 2020, MNRAS, 494, 3712
Line J. L. B., Webster R. L., Pindor B., Mitchell D. A., Trott C. M., 2017, PASA, 34, e003
Liu A., Shaw J. R., 2020, PASP, 132, 062001
Liu A., Tegmark M., 2011, Phys. Rev. D, 83, 103006
Liu A., Parsons A. R., Trott C. M., 2014a, Phys. Rev. D, 90, 023018
Liu A., Parsons A. R., Trott C. M., 2014b, Phys. Rev. D, 90, 023019
Lomb N. R., 1976, Ap&SS, 39, 447
Masui K. W. et al., 2013, ApJ, 763, L20
Mertens F. G. et al., 2020, MNRAS, 493, 1662
Morales M. F., Hewitt J., 2004, ApJ, 615, 7
Morales M. F., Hazelton B., Sullivan I., Beardsley A., 2012, ApJ, 752, 137
Neben A. R. et al., 2015, Radio Sci., 50, 614
Neben A. R. et al., 2016, ApJ, 826, 199
Newburgh L. B., et al., Proc. SPIE 9906, Ground-based and Airborne Telescopes VI, 2016, , SPIE, Bellingham, WA. p. 99065X
Parsons A., Pober J., McQuinn M., Jacobs D., Aguirre J., 2012a, ApJ, 753, 81
Parsons A. R., Pober J. C., Aguirre J. E., Carilli C. L., Jacobs D. C., Moore D. F., 2012b, ApJ, 756, 165
Parsons A. R. et al., 2014, ApJ, 788, 106
Patil A. H. et al., 2016, MNRAS, 463, 4317
Patra N. et al., 2018, Exp. Astron., 45, 177
Pober J. C. et al., 2012, AJ, 143, 53
Pober J. C. et al., 2013, ApJ, 768, L36
Pober J. C. et al., 2014, ApJ, 782, 66
Scargle J. D., 1982, ApJ, 263, 835
Shaw J. R., Sigurdson K., Pen U.-L., Stebbins A., Sitwell M., 2014, ApJ, 781, 57
Slepian D., 1978, Bell Syst. Tech. J., 57, 1371
Solomon O. M. J., 1993, NASA STI/Recon Technical Report N, IEEE, New York, NY
Subrahmanya C. R., Manoharan P. K., Chengalur J. N., 2017, J. Astrophys. Astron., 38, 10
Switzer E. R. et al., 2013, MNRAS, 434, L46
Switzer E. R., Chang T. C., Masui K. W., Pen U. L., Voytek T. C., 2015, ApJ, 815, 51
Tegmark M., 1997, Phys. Rev. D, 55, 5895
Thompson A. R., Moran J. M., Swenson, George W. J., 2017, Interferometry and Synthesis in Radio Astronomy, Springer, London, UK
Thyagarajan N., Parsons A. R., DeBoer D. R., Bowman J. D., Ewall-Wice A. M., Neben A. R., Patra N., 2016, ApJ, 825, 9
Tingay S. J. et al., 2013, PASA, 30, e007
Trott C. M. et al., 2016, ApJ, 818, 139
van Haarlem M. P. et al., 2013, A&A, 556, A2
Vedantham H., Udaya Shankar N., Subrahmanyan R., 2012, ApJ, 745, 176
Virtanen P. et al., 2020, Nat. Methods, 17, 261
Zhang Y. G., Liu A., Parsons A. R., 2018, ApJ, 852, 110
Zheng H. et al., 2017, MNRAS, 464, 3486

APPENDIX A: THE DEPENDENCE OF CLEAN RESIDUAL AMPLITUDES ON THE TOLERANCE PARAMETER

In our comparison, we assumed a fixed set of CLEAN parameters employed by the HERA pipeline (Kern et al. 2019) and the RFI environment of the Karoo radio observatory. The presence of flagging leaks residuals left over by CLEANing across all delays. Hampering a 21-cm detection. Lowering the residuals also lowers this leakage so in principal decreasing the tolerance should allow for sufficiently low residuals for a 21-cm detection. In this appendix, we examine the CLEAN performance as a function of flagging percentage and tolerance parameter. We run CLEAN for a single model baseline and time across all 256 channels with 256 channel zero-padding on either side and a Tukey taper. We iteratively increase the width of flagging on the ORBCOMM band; starting with no flags, then introducing two 235 kHz channels centred at 137 MHz. Next, we introduce four channels, eights channels, and 16 channels. In the top-panel of Fig. A1, we compare residuals for different levels of flagging to the injected 21-cm signal. Even when two channels are flagged, significant deviations are introduced in CLEAN when the tolerance is set to 10^{-9} (solid coloured lines). On the other hand, DAYENUREST reproduces both the foregrounds and signal with no residual bias.

As we mentioned above, the biases from CLEAN arise from foreground residuals that have not been fully subtracted and still contain sidelobes from flagging. By decreasing the tol parameter in CLEAN, we can actually subtract deeper. Thus, in principal there should exist small enough values of the tolerance such that sidelobes are suppressed enough to recover 21 cm fluctuations without significant foreground bias. We explore this possibility by lowering the tolerance to 10^{-11} (Fig. A1 bottom panel). Given this lower value, residuals

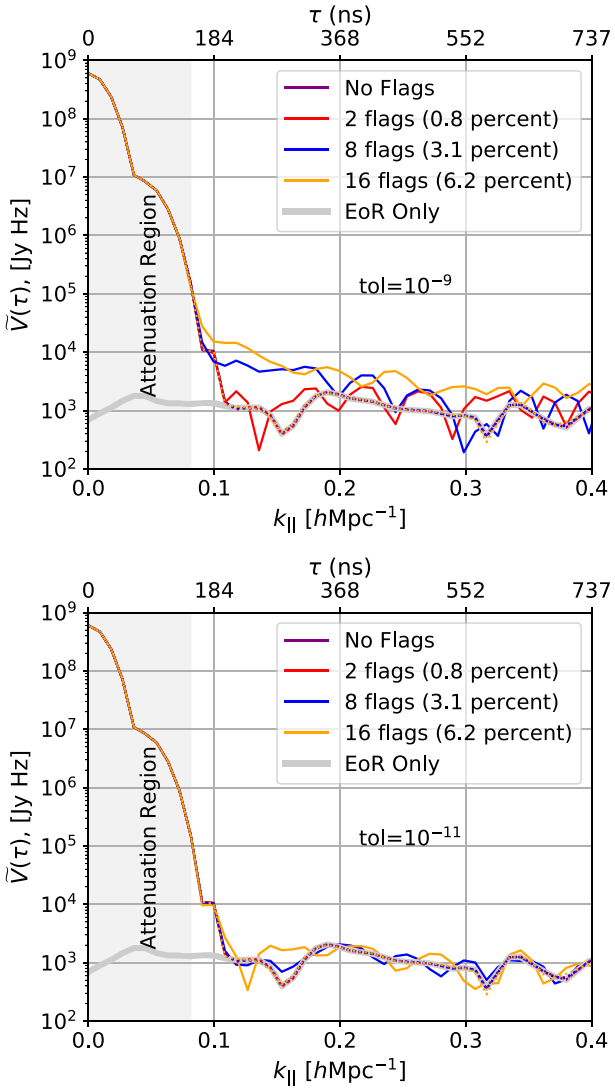

Figure A1. Top: Delay-transformed CLEANed visibilities for $\texttt{tol} = 10^{-9}$ (top panel) and $\texttt{tol} = 10^{-11}$ (bottom panel). Different colours denote different numbers of contiguous flagged channels centred at the 137 MHz ORBCOMM frequency. No other flags are introduced and CLEAN is performed over the entire band. Dotted lines are the results of applying DAYENU to the various levels of flagging. The DAYENU filtered visibilities are in very good agreement with the signal outside of the attenuation region.

are not visibly present with two flagged channels but $\gtrsim$10 per cent biases appear after $\gtrsim$8 channels (only 3.1 per cent of the data) are flagged. Running CLEAN with $\texttt{tol} = 10^{-11}$ takes 22 s per baseline and time-sample on a 2.4 GHz i5 processor – ∼100 times slower then the linear filter if $\mathbf{R}^{\neg}$ is computed at every baseline time and $\sim 10^4$ times slower then the realistic scenario where all baseline-times can be filtered with cached matrices.

While decreasing the tolerance can lower foreground leakage, there are diminishing returns and even after a 10^4 performance hit relative to DAYENU, we run into trouble with just 3 per cent of channels flagged.

This paper has been typeset from a TEX/LATEX file prepared by the author.